NSU Florida
NOVA SOUTHEASTERN UNIVERSITY

Nova Southeastern University
**NSUWorks**

HCNSO Student Theses and Dissertations | HCNSO Student Work

10-28-2015

# Fine-Grained Bacterial Compositional Analsysis of the Port Everglades Inlet (Broward County, FL) Microbiome using High Throughput DNA Sequencing

Lauren M. O'Connell
*Nova Southeastern University*, lo248@nova.edu

Follow this and additional works at: https://nsuworks.nova.edu/occ_stuetd

Part of the Marine Biology Commons, and the Oceanography and Atmospheric Sciences and Meteorology Commons

## Share Feedback About This Item

This Thesis has supplementary content. View the full record on NSUWorks here: https://nsuworks.nova.edu/occ_stuetd/391

NSUWorks Citation

Lauren M. O'Connell. 2015. *Fine-Grained Bacterial Compositional Analsysis of the Port Everglades Inlet (Broward County, FL) Microbiome using High Throughput DNA Sequencing*. Master's thesis. Nova Southeastern University. Retrieved from NSUWorks, . (391) https://nsuworks.nova.edu/occ_stuetd/391.

This Thesis is brought to you by the HCNSO Student Work at NSUWorks. It has been accepted for inclusion in HCNSO Student Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

HALMOS COLLEGE OF NATURAL SCIENCES AND OCEANOGRAPHY

# Fine-Grained Bacterial Compositional Analysis of the Port Everglades Inlet (Broward County, FL) Microbiome using High Throughput DNA Sequencing

By

Lauren Marie O'Connell

Submitted to the Faculty of
Halmos College of Natural Sciences and Oceanography
in partial fulfillment of the requirements for
the degree of Master of Science with a specialty in:

Marine Biology
and
Coastal Zone Management

Nova Southeastern University

October 2015

# Thesis of
# Lauren O'Connell

Submitted in Partial Fulfillment of the Requirements for the Degree of

# Masters of Science:

# Marine Biology and Coastal Zone Management

Lauren O'Connell
Nova Southeastern University
Halmos College of Natural Sciences and Oceanography

October 2015

Approved:

Thesis Committee

Major Professor :______________________________
Jose Lopez, Ph.D.

Committee Member :___________________________
Song Gao, Ph.D.

Committee Member :___________________________
Donald McCorquodale, Ph.D.

**TABLE OF CONTENTS**

ACKNOWLEGEMENTS 5
ABSTRACT 6
List of Figures 7
List of Tables 7
INTRODUCTION 8
History of Marine Microbiology 8
The Role of Microbes in the Ocean 12
Ribosomal RNA and the 16S Subunit 14
High-throughput DNA Sequencing 15
Metagenomics/Microbiomes and 16S Amplicon Sequencing 18
Land-based Pollution Sources and their Impacts in Southeast Florida 19
Florida's Ports and Port Everglades Inlet 21
HYPOTHESES AND OBJECTIVES 23
METHODS 24
Water Sample Collection and Filtration 24
DNA Extraction 25
Ion Chromatography (For Water Chemistry Analysis) 26
Illumina Miseq High-throughput Sequencing Sample Preparation 27
Sequencing Analysis: QIIME Analysis 29
Sequencing/Statistical Analysis: QIIME and R Statistical Software 30
Importing, Filtering, and Normalization 30
Alpha Diversity QIIME 31
Alpha Diversity Phyloseq 31
Beta Diversity 31
Pathogen Detection 32
Regression Analysis with Environmental Metadata 32
RESULTS 33
MiSeq Output 33
Alpha Diversity 34
Rarefaction Curves 34
Rank Abundance Curves 38
Alpha Diversity Indices 39
Relative Abundance of Microbes in Port Everglades Inlet 40
Beta Diversity 45
Location 45
Season 47
Month 48
Pathogens 53
*Bacillales, Clostridiales,* and *Lactobacillales* 53
*Enterobacteriales, Campylobacterales,* and *Vibrionales* 55
Ion Chromatography and Environmental Metadata Correlations, and Linear Regression Analysis 59

DISCUSSION 60
Port Everglades Inlet and Sample Collection 61

Alpha Rarefaction Curves as a Measure of Sampling and Sequencing Efforts 62
Rank Abundance Curve 63
Comparison of Alpha Diversity Indices 63
Microbial Community Composition Taxa Fluctuations: Location, Month, and Season 64

Significance of Abundant Taxa and Correlation with Environmental Metadata 65
*Alphaproteobacteria* 66
*Gammaproteobacteria* 67
*Betaproteobacteria* 69
*Flavobacteria* 70
*Chloroplast* 71
*Synechoccophycideae* 72
*Acidomicrobiales* 72
*Planctomycetia* 73
*Verrucomicrobiae* 73
Presence of Pathogens in Port Everglades Inlet Surface Waters 74
*Bacillales* 75
*Clostridia* 76
*Campylobacteriales* 77
*Lactobacillus* 78
*Enterobacteriales* 79
*Vibrionales* 79
Overall Presence of Pathogens in Port Everglades Inlet 79
Comparison of Illumina Data with 454 Pyrosequencing 92
IMPORTANCE AND SIGNIFICANCE 80
CONCLUSIONS 81
REFERENCES 84
APPENDICES 96

**Acknowledgements**

I thank my major advisor Jose Lopez for the support and guidance he has provided me throughout my thesis research. I thank my thesis committee members and associated faculty members: Song Gao, Donald McCorqudale, and Jay Fleisher for their support and expertise throughout my thesis research, statistics, and writing. I also thank Nova Southeastern University for the President's Grant that provided me with financial support to complete my thesis research. I thank my fellow lab mates and friends who helped me with my sample collections, as well as providing encouragement and support throughout my thesis project: Amber Metallo, Emily Powell, Josh Stocker, Andrew Calhoun, Alex Campbell, and Renee Potens. I would like to extend a special thank you to my family, John, Linda, and Jackie O'Connell who have always encouraged me to follow my dreams and continue on with my education, even when I felt like quitting. Finally, I thank my undergraduate professors: Dr. Ryan, Dr. Aruscavage, Dr. Oakley, and Dr. Junker for giving me a strong background in the marine sciences and for helping me to realize my passion for marine microbiology.

**Abstract:**

Port Everglades Inlet is one of the busiest ports in the country and is a point source of pollution to surrounding beaches and offshore corals from heavy boat traffic and urban runoff. Understanding fluctuations of bacterioplankton communities in major port inlets is important due to their impacts on surrounding marine environments. To understand annual microbial fluctuations, the 16s rRNA V4 hypervariable region was sequenced using Illumina high-throughput DNA sequencing technology. Surface samples were taken weekly for one year to generate baseline fluctuations in the microbial community. Total reads of 1.4 million were generated with a final count of 16,384 Operational Taxonomic Units. The dominant phyla were Proteobacteria, Cyanobacteria, Bacteroidetes, and Actinobacteria. Pathogenic genera were detected at low abundances during peak shipping and tourist months (November –April). Results indicate significant differences in alpha diversity when comparing microbial communities in August with other times. This was likely caused by low community richness and abundance, and below-average August rainfall levels. Differences in beta diversity were significant when comparing monthly and seasonal changes. Rainfall, temperature, and nutrient trends may have affected microbial composition, specifically during the dry season that was warmer and wetter than historical averages for 2013-2014. Increased nitrogen and phosphorous concentrations were observed in the dry season months of October, December, and January potentially creating optimal bacterial growth conditions. These results can be compared with historical and future data regarding inlet microbial communities to determine underlying baselines of bacterioplankton communities and monitor the health of marine and recreational environments they impact. This study represents the first to characterize at this scale and use Illumina MiSeq technology to analyze water samples from Port Everglades.

**Keywords:** *Microbiome; 16S; Ribosomal RNA; Bacterioplankton; Port Everglades; Inlet; Illumina*

## List of Figures

**Figure 1:** Port Everglades Inlet sampling locations.
**Figure 2:** Workflow for Illumina MiSeq 16s rRNA sample preparation.
**Figure 3:** Rarefaction curves of month for Chao1 diversity indices generated in QIIME.
**Figure 4:** Rarefaction curves of month for Shannon diversity indices generated in QIIME.
**Figure 5:** Rarefaction curves of month for Observed Species and Shannon diversity indices generated in R Statistical Software.
**Figure 6:** Rank abundance plot of top 500 OTUs in dataset.
**Figure 7:** Community richness boxplot by month.
**Figure 8:** Alpha diversity index plots.
**Figure 9:** Average relative abundance vs. frequency plot for top 6 most abundance bacterial phyla in dataset.
**Figure 10:** Taxa summary of relative abundance at phylum taxonomic level.
**Figure 11:** Taxa summary of relative abundance at class taxonomic level.
**Figure 12:** Taxa summary of relative abundance at genus taxonomic level.
**Figure 13:** NMDS plot of Bray-Curtis dissimilarity by location.
**Figure 14:** NMDS plot of weighted Unifrac distance by location.
**Figure 15:** NMDS plot of Bray-Curtis dissimilarity by season.
**Figure 16:** NMDS plot of weighted Unifrac distance by season.
**Figure 17:** NMDS plot of Bray-Curtis dissimilarity by month.
**Figure 18:** NMDS plot of weighted Unifrac distance by month.
**Figure 19:** Taxa summary and abundance levels of top 50 OTUs in Bacillales order.
**Figure 20:** Taxa summary and abundance levels of top 50 OTUs in Clostridiales order.
**Figure 21:** Taxa summary and abundance levels of top 50 OTUs in Lactobacillales order.
**Figure 22:** Taxa summary and abundance levels of top 50 OTUs in Enterobacteriales order.
**Figure 23:** Taxa summary and abundance levels of top 50 OTUs in Campylobacteriales order.
**Figure 24:** Taxa summary and abundance levels of top 50 OTUs in Vibrionales order.
**Figure 25:** Correlation between top phyla abundance and environmental metadata: salinity, temperature, and precipitation

## List of Tables

**Table 1:** Latitude and longitude of sites where water samples were collected.
**Table 2:** Results of multiple comparison Adonis tests by month.
**Table 3:** Results of multiple least squares linear regression analysis using SAS.

**Introduction:**

Marine microbes are the dominant organisms in the world's oceans contributing over 98% of the biomass. Without marine microbes many of the major biogeochemical cycles are not possible, and the planet would cease to function in the ways observed every day. Marine microbes impact humans, animals, and plants both positively through nutrient availability, and negatively through their ability to cause sickness. Yet, very little is known about majority of the roles they play in the environment. There is even less information on the community structures of marine microbes.

This study aims to complete a microbiome characterization of the surface water in Port Everglades Inlet located in Broward County, FL. Water samples were collected from June 2013 – May 2014 to determine monthly alpha (α) and beta (β) diversity fluctuations. Alpha diversity is the diversity within an ecosystem or sample and is often expressed in terms of species richness (Whittaker, 1972). An ecosystem with high alpha diversity will have a high number of species with similar abundances. Beta diversity compares diversity between ecosystems showing change or differentiation in the species composition between samples. (Whittaker, 1972; Koleff *et al.,* 2003). This study examines changes in composition of Port Everglades Inlet's surface water microbiome over a year. This study differs from others because samples were taken on a weekly basis allowing for changes to be examined at a finer scale (or higher time resolution). This study also uses Illumina 's MiSeq high-throughput sequencing technology to complete DNA sequencing of water samples, differing from previous studies which were largely restricted to culture-based methods.

**History of marine microbiology:**

Marine microbes were not known to play an important role in marine ecosystems before the late 1970's (Pedros-Alio, 2006). As of 1974, studies completed on relative abundance of microbes in the sea were highly contradictory (Pomeroy, 1974). Studies by Kris (1963) and Seki (1972) found substantial numbers of microbes in their samples, while contradictory studies by Wiebe and Pomeroy found trivial amounts (1972). The

major cause of incongruities within these studies was due to the use of different methods for bacterial isolation. Coordination of methods between scientists was uncommon, resulting in highly conflicted microbial abundance counts across different studies (Pomeroy, 1974).

Development of nucleopore filters and epifluorescent microscopy in the late 1970's allowed for accurate approximations of the total number of microbes in the sea (DeLong, 2009; Hobbie *et al,* 1977; DeLong and Karl, 2005). Prior to the use of nucleopore filters, cellulose filters were used. Cellulose filters have rough and uneven surfaces, resulting in inaccurate cell counts because bacteria become trapped within the filter instead of on the observable surface. Subsequently, these bacteria are not included in the cell counts when using epifluorescent microscopy. Nucleopore filters differ from cellulose filters because they have a homogenous pore size and a very flat surface. This ensures that all bacteria remain on the observable surface of the filter and are included in the microbial abundance counts (Hobbie *et al,* 1977). Development of these technologies revealed that many of the studies completed on marine microbes and marine ecology grossly underestimated the amount of bacteria present in the world's oceans (DeLong and Karl, 2005). This is known as the "great plate count anomaly," where scientists noticed considerable inconsistencies in their plate counts versus their microscopy counts (Handelsman, 2004). It was hypothesized that some studies underestimated the number of microbes present in a milliliter of seawater by at least three orders of magnitude (DeLong and Karl, 2005).

In 1978 another important discovery in marine biology propelled the field of marine microbiology forward. It was in this year that the deep-sea submersible *Alvin* explored the hydrothermal vent communities along the Galapagos Rift (Corliss *et al,* 1979). This study changed the way that most scientists looked at deep-sea environments. Prior to this, the deep sea was believed to be a barren, desert-like ecosystem (Gaille, 1993). This study reported the presence of large concentrations of sulfur-oxidizing bacteria (chemoautotrophs), heterotrophic bacteria, and invertebrates present at the hydrothermal vents (Corliss *et al,* 1979 and DeLong and Karl, 2005). Studies completed on the chemoautotrophic bacteria after the *Alvin* exploration revealed that invertebrates

present at the vents received their nutrition through chemoautotrophic bacterial symbiosis (Felbeck, 1981; Cavanaugh *et al,* 1981; Gaille, 1993). This discovery confounded the scientific community by revealing that marine microbes are present everywhere in the ocean, and that they play a major role in the food chain at the vent communities.

Since the discovery of hydrothermal vent bacteria, many new technologies have been developed to study marine microbes and their functions in marine environments (DeLong and Karl, 2005). The discovery of viable but nonculturable (VBNC) microbes in 1981 pressured scientists to discover new mechanisms to identify bacteria without using culture-dependant methods (Xu *et al,* 1982; Handelsman, 2004). Sequencing of microbial rRNA and Polymerase Chain Reactions (PCR) are two technologies developed to overcome the problem of viable but nonculturable bacteria (Giovannoni *et al,* 1990). The establishment of known rRNA sequences from hydrothermal vent bacteria enabled microbes to be compared at the macromolecular level (DeLong and Karl, 2005). Insights into the phylogenetic relationships of microbes were now available without having to culture them (Stahl *et al,* 1984). The use of PCR in marine microbial ecology and biology began in the early 90's (Giovannoni *et al,* 1990). The advantage of using PCR for nonculturable microbes is that a specific portion of the DNA is targeted and then amplified, generating millions of copies of a specific piece of DNA (Saiki *et al.* 1988). This overcomes limitations when DNA is sparse or when there is nonessential portions of DNA present in data (Randall *et al,* 1988). A study in the Sargasso Sea by Giovannoni and colleagues used PCR to amplify 16S rRNA genes of natural populations (1990). They found novel microbial groups that differed from any that had been previously cultured (Giovannoni *et al,*1990). This study demonstrates the capacity to which PCR can enhance research efforts in marine microbiology without the need to use cultivation techniques.

The field of marine microbiology has had many major breakthroughs since discovering microbes in the oceans. Developments in genome science and technology allow for expansive research opportunities in marine microbiology and ecology (Handelsman, 2004, DeLong and Karl, 2005). The first marine archaeal organism to successfully have its genome fully sequenced is *Methanocaldococcus jannaschii* (Bult *et*

*al,* 1996). Bult and colleagues discovered that only 38 percent of the genes found in the*M. jannaschii* genome had known functions (1996). The genes involved in metabolic processes are more similar to bacterial genes, while genes involved in transcription, translation, and replication are more similar to eukaryotic genes (Bult *et al,* 1996). In 2003, the genomes of multiple strains of the cyanobacteria. *Prochlorococcus,* were sequenced (Rocap *et al,* 2003, and Dufresne *et al,* 2003). *Prochlorococcus* is the first marine bacterial species to have its genome fully sequenced (Whitfield, 2003). Up to 2006, over 150 marine microbial genomes have been sequenced. From this data scientists discovered more about the evolution of specific genes and genomes, and the effects that an organism's genome has on its ecological niche and ecosystem functions (Pedros-Alio, 2006). Due to decreased costs of sequencing technology, there has been a rapid increase in the number of microbial genomes that are being sequenced. To organize the rapid influx of metagenome and full genome sequences the Genomes OnLine Database (GOLD) was created. The GOLD database is available on the Internet and provides access to all genome and metagenome projects and their associated metadata worldwide. To date, the total number of sequencing projects completed, or in progress on the GOLD database is 63,489 with greater than 40,000 of those projects being completed on microbial species (Pagani *et al,* 2012 and Reddy *et al,* 2014).

The past 40 years has been an era of discovery and growth in the field of marine microbiology. Technological advances allowed accurate estimations of microbial populations, sequencing of rRNA/DNA amplicons, and now, full genome sequencing. From these new technologies, many old questions posed about microbial population ecology, evolution, and biology, can be reassessed using a genomic approach. Large databases of known and unknown genes are being created, which contributes information to microbial gene evolution and how this impacts microbial ecological niches and community structures. There is still much to be learned about marine microbiology, but with the application of techniques such as culturing, PCR, DNA/RNA amplicon sequencing, and genome sequencing large magnitudes of information can be gained.

**The Role of Microbes in the Ocean:**

Marine microbial communities play a large role in the oxidation of organic compounds, and nutrient regeneration in the marine environment (Herbert, 1999). They are major components in global biogeochemical cycles, especially carbon, nitrogen and phosphorous cycles (Azam et al 1983; Arrigo, 2005). Nitrogen is one of the most abundant elements on the planet, and is essential for the production of nucleic acids. The nitrogen cycle is a series of redox reactions that transforms nitrogen gas ($N_2$) to ammonium ($NH_4^+$) (Canfield *et al,* 2010). $N_2$-fixing bacteria are common in light abundant, tropical and subtropical waters (Arrigo, 2005). Inorganic nitrogen is often depleted in these areas due to high primary productivity (Arrigo, 2005; Brandes *et al,* 2007). Fixing of nitrogen gas by bacteria allows it to be converted into inorganic nitrogen, which is the form most often utilized by other organisms, especially nitrifying bacteria (Gruber, 2004 and Arrigo, 2005). Nitrogen fixation must be in balance with the total amount of fixed nitrogen lost to processes such as denitrification. The dominant $N_2$-fixing microbe in the oceans is the cyanobacterium, *Trichodesmium.* The rate at which *Trichodesmium* fixes $N_2$ is estimated in the range of 22- 34 Tg N $yr^{-1}$, and accounts for roughly 36-50% of the nitrogen demands of the marine microbial community (Arrigo, 2005). Denitrification is the process in which reduction of nitrogen to dinitrogen occurs (Alvarez, *et al,* 2014). It can occur in both anaerobic and aerobic environments, and utilizes nitrate as the terminal electron acceptor instead of oxygen (Brandes *et al,* 2007). A major denitrifier in the oceans is *Planctomycetes,* a facultative aerobe or anaerobe. This group is chemoorganotrophic and is found in polluted waters (Siniscalchi *et al,* 2015). Some species within the *Planctomycetes* genus are capable of completing anammox reactions in which ammonium is oxidized to $N_2$ (Arrigo, 2005; Brandes *et al,* 2007). Availability of fixed nitrogen in the form of nitrate ($NO_3^-$) is the major limiting factor for biological production in the mixed surface layer (Gruber, 2004). Research has shown that marine $N_2$-fixation is more important than previously thought for primary production, especially in tropical and sub-tropical environments (Falkowski, 1997; Gruber, 2004). Estimates indicate that the largest amount of nitrogen fixation occurs in pelagic marine environments at a rate of 117 Tg N $yr^{-1}$ (Codispoti *et al,* 2001). Marine nitrogen-fixing bacteria are responsible for the transformation of $N_2$ into $NO_3^-$,

maintaining the balance of biologically available nitrogen, and are therefore of unconditional importance to nutrient cycling between the atmosphere and the world's oceans (Canfield *et al*, 2010).

The phosphorus cycle is another major biogeochemical cycle in the oceans. Phosphorous is a limiting, yet vital nutrient in the oceans. It is necessary for primary production in oceanic environments and is needed for the survival of marine microbes and phytoplankton (Froelich *et al,* 1982). Without the availability of usable phosphorus, primary production is limited, negatively impacting phytoplankton and bacterial populations, and reducing the rate of phosphorous cycling (Dyhrman *et al,* 2007).

The main form of phosphorus biologically available for most marine organisms is dissolved inorganic phosphorus (DIP), but studies show that dissolved organic phosphorous (DOP) is present in levels similar to those of DIP, and is readily taken up by plants (Johannes, 1964). Phosphorus availability in the marine environment is believed to be a major driver of evolutionary adaptations in marine microbes and phytoplankton to increase their rates of phosphorous cycling. In a previous study it is shown that microbial community composition changes in response to availability of nutrients in their environment. When phosphorous concentrations are low major bacterial species are absent (Pinhassi *et al,* 2006). Organisms that developed adaptations for low phosphorus environments are *Prochlorococcus* and *Synechococcus*. Synthesis of genomic DNA requires a large phosphorous demand. These two bacterial groups have some of the smallest genome sizes known. This is a possible adaptation developed due to their presence in low phosphorous environments. A second adaptation developed by these two bacterial groups is that they can synthesize lipids composed of sulfur and sugar rather than lipids formed from phosphate (Dyhrman *et al,* 2007). These adaptations make them suitable and successful in low phosphorous environments, where other bacterial species cannot persist.

**Ribosomal RNA and the 16S subunit:**

The ribosomal RNA (rRNA) molecule is generally accepted as a universal and comparative molecule for bacterial phylogeny and taxonomy due to major characteristics outlined by Woese and colleagues (1980,1982,1984, and 1985). rRNA is present in almost all bacteria and is part of a large complex that is vital for cell function. The largeness and important function of this molecule, paired with evidence that its function has been conserved over time and not transferred from species to species allows for reliable phylogenetic comparisons between organisms using rRNA (Woese *et al,* 1980, 1982, 1984, and 1985, Janda and Abbott, 2007). The prokaryotic rRNA molecule is composed of two subunits, the large and the small. 16S rRNA is part of the small subunit of prokaryotic ribosomes. The genes that encode this region are called 16S rDNA and are used to determine taxonomy of microbes. The 16S gene is ideal for taxonomy studies because it is highly conserved allowing for broad taxonomic comparisons between organisms (Woese, 1987). The 16S gene is most commonly used for taxonomic purposes in sequencing studies because the full gene is not being sequenced. This impacts the discriminatory power at the genus level and gives low phylogenetic power at the species level. Only a full 16S gene sequence provides accurate phylogenetic comparisons between species (Janda and Abbott, 2007 and Birtel *et al,* 2015). The 16S gene contains 9 hypervariable regions (V1-V9), each with different evolutionary rates and sequence diversity (Tringe, 2009). The hypervariable regions with rapid genetic changes help to determine evolutionary changes between species from the same genus leading to more accurate classifications (Chakravorty *et al,* 2007). There is not a single hypervariable region that is used to distinguish between all bacteria, and therefore the development of primers designed to amplify one or more hypervariable regions are used. There is disagreement in the scientific community about which hypervariable regions give the most accurate classification results. For example, Wang *et al.*(2007) claim that the V2/V4 regions yield the most accurate results, while Chakravorty *et al.*(2007) claim it is the V2/V3 regions, while Sundquist *et al.*(2007) claim it is the V2/V3/V4 regions. Ultimately, the ideal hypervariable region to be used for 16S gene taxonomy varies with the sequencing technology that is used. In older high-throughput sequencing technologies (454 pyrosequencing), the V6 region is used because of its short read length.

Development of Illumina high-throughput sequencing now allows for longer read lengths, and studies show that the V3-V4 regions generate the best taxonomic results (Shah *et al,* 2010, He *et al,* 2013 and Fadrosh *et al,* 2014). Overall, it is noted that hypervariable regions with longer basepair (bp) lengths and semiconserved, rather than conserved regions surrounding them, increase the accuracy of bacterial taxonomic classifications (Wang *et al,* 2007).

**High-throughput DNA sequencing**

Methods for DNA sequencing have been developed and re-developed since the late 1970's. Maxam and Gilbert developed one of the first DNA sequencing methods in 1977 called Maxam-Gilbert sequencing or chemical sequencing. They established a method where chemical processes are used to break apart a DNA molecule into its nucleotide bases called chemical degradation (Maxam and Gilbert, 1977, Sanger *et al,* 1977). The DNA is terminally labeled and produces radioactive DNA fragments that are run on a polyacrylamide gel. The DNA sequence is then read from the patterns observed on the gel (Maxam and Gilbert, 1977). While this provided DNA sequences, there are some limitations to this technique (Schuster, 2008). The resolution of the polyacrylamide gel is not always sufficient for analysis and a maximum of only 100 base pairs (bp) can be sequenced (Maxam and Gilbert, 1977). Later, in that same year, Sanger and Coulson developed a methodology for rapid DNA sequencing that improved upon previous techniques. This method is called "Sanger sequencing" or dideoxynucleotide sequencing. Sanger sequencing utilizes DNA polymerase and 2',3'- dideoxythymidine triphosphates(ddTTPs) to terminate DNA chain elongation (Sanger *et al,* 1977). This termination occurs at places in the DNA where thymidylic acid(dT) would be incorporated. The chain elongation terminates at this spot because ddTs do not contain a 3' hydroxyl group and therefore cannot form a phosphodiester bond with the next base. In Sanger sequencing a mixture is made containing commercially available DNA polymerase, a mixture of ddTTPs, and all the dTTPs (dATP, dCTP, dGTP, dTTP). The mixture is made so that the concentration of the ddTTP is only 1% of he total dTTP concentration, this way the chain will terminate at whichever ddTTP is added (Sanger *et al,* 1977). This reaction is completed 4 times, once for each base. The reactions are run on

thin polyacrylamide gels and analyzed by size of the fragment (Sanger *et al,* 1977 and Sanger and Coulson, 1978). This was the main method of DNA sequencing throughout the next three decades (Schuster, 2008; Metzker, 2010). It was during this time, scientists began to anticipate the possibility of sequencing the entire human genome (Schuster, 2008). In 1990 the Human Genome Project (HGP) was initiated (Venter *et al,* 2001). High-throughput sequencing was developed due to the eagerness of scientists to sequence the entire human genome (Schuster, 2008 and Venter *et al,* 2001). The first complete human genome was sequenced in April of 2003 (add a citation).

After the HGP concluded, scientists were interested in continuing research using next-generation sequencing technologies. This led to the development of four main high-throughput DNA sequencing platforms. These platforms are PacBio, Life/SOLiD, 454/ Roche-Pyrosequencing, and Illumina/Solexa (Shendure and Ji, 2008). High-throughput DNA technology allows for Megabase-sized quantities of DNA to be sequenced in just a few hours, and at a relatively cheap price (Quail *et al,* 2008 and Metzker, 2010).

454 pyrosequencing is the first next generation sequencing platform that was made commercially available (Margulies *et al,* 2005; Shendure and Ji, 2008). It works through the use of emulsion PCR. The genome is split into fragments and the fragments are attached to beads, which are added into an oil emulsion containing PCR reactants (Margulies *et al,* 2005 and Mardis, 2011). This process allows for bulk amplification of the DNA fragment, and eliminates the need for cloning in bacteria (Mardis, 2011). Sequencing begins with the amplified beads being incubated with *B. stearothermophilus* polymerase and binding proteins in small wells. This makes them biochemically compatible with the sequencing technique (Shendure and Ji, 2008). Extra, smaller beads are added to the mix, which contains other enzymes needed for the pyrosequencing reaction. The sequencing occurs in cycles, with each cycle having different species of unlabeled nucleotide being washed over the microplate array (Shendure and Ji, 2008). When the nucleotide is incorporated into the sequence, a release of inorganic pyrophosphate occurs, causing a burst of light (Margulies *et al,* 2005; Shendure and Ji, 2008; Mardis, 2011). This burst of light is detected by a charge-coupled device (Margulies *et al,* 2005; Shendure and Ji, 2008). The intensity of the light reaction

indicates how many of the same nucleotides are present in a row (Shendure and Ji, 2008). The benefits of pyrosequencing is that it gives relatively long read lengths of roughly 250 bp, and the time it takes to run is relatively short, around 7 hours. The negatives of 454 pyrosequencing are that it has a high homopolymer error, and is more expensive per base when compared to Illumina/Solexa (Mardis, 2008; Metzker, 2010).

Illumina/Solexa is now the most commonly used high-throughput DNA sequencing platform due to its lower cost per sequence (Metzker, 2010; Caporaso *et al,* 2012). The Illumina platform is used in multiple microbiome studies to gain insight into the complex diversity present within many different environments (Gilbert *et al,* 2010; Caporaso *et al,* 2012; Kelley and Gilbert, 2013) Illumina sequencing technology is also used in numerous coastal marine microbiome studies, providing an accurate overview of the complex microbial communities (Gilbert *et al,* 2010). This platform works by first ligating adapters to target DNA fragments (Mardis, 2008; Mardis, 2011). The fragments with attached adaptors are bound to a glass flow cell containing 1 or more channels (Mardis, 2011). Unlabeled nucleotides and enzymes are added to the chambers to initiate bridge amplification of the DNA fragments. The bridge amplification step creates dense clusters of the DNA fragments on the flow cell. Sequencing occurs by single base extension and is completed by adding fluorescently labeled reversible terminator nucleotides, primers, and DNA polymerase. The fluorescent signal emitted from the clusters is recorded by a camera and the nucleotide base is determined. The blocked 3' end and the fluorescent tag are removed from the end of the fragment and a new cycle begins (Mardis, 2008; Mardis, 2011). The major benefits of Illumina sequencing are that it is very cheap per base and the chemistry behind the reactions is very advanced (Metzker, 2010).

**Metagenomics/Microbiomes and 16S Amplicon Sequencing**

Advancements in DNA sequencing have equipped scientists with tools to examine microbial environmental communities (microbiomes) with less bias yet greater coverage and depth. Before the advances in molecular techniques many studies worked only with those micro-organisms that could be cultured (Handelsman, 2004). The term "metagenomics" was first coined by Handelsman *et al.*in 1998, and is defined as a molecular biology approach, absent of culture techniques, which aims to complete genetic, taxonomic, and functional analyses of cumulative microbial genomes comprising a specific microbiome. The original approaches to analyzing metagenomes relied heavily on PCR amplification and gene cloning of genes with sequence similarity. The clones obtained were used to reconstruct genomes by aligning overlapping fragments and determining the sequence of the chromosome clone by clone (Schloss and Handelsman, 2003). A study by Rondon and colleagues to used PCR amplification and gene cloning to examine the soil microbiome (2000). This study used bacterial artificial chromosomes (BACs) from pBeloBAC11 as their vector, which was used to generate clones for sequencing. This study is important because it demonstrates that modern molecular biology methods could be used to generate data about a microbial population using non-culture techniques.

Ventor and colleagues(2004) completed a revolutionary oceanic microbiome study in the Sargasso Sea examining the surface water microbiome. This study used whole-genome shotgun sequencing to obtain information about microbial genomes. This approach is similar to cloning methods used previously, but PCR amplification is not needed. Whole-genome shotgun sequencing is used because of the tendency for PCR reactions to be biased when working with 16S rRNA. Not all 16S rRNA regions amplify with the same universal primers, leading to some regions being over-represented and others, under-represented. While this study was confined to a relatively small area of the planet, it provided many important insights into the microbial community composition and its function in the Sargasso Sea. One of the most interesting findings in this study was the possible mechanism by which microbial organisms utilize the available phosphorus in an exceptionally low phosphorous environment. Genes were found that are

known to aid in the transport mechanism of phosphonates (Venter *et al,* 2004). This mechanism was previously identified in both *Prochlorococcus* and *Synechoccus* genomes, the latter being one of the major microbial groups identified in the sequencing data from this study (Venter *et al,* 2004). When this study was conducted in 2004, the price of sequencing to such great depths was too expensive for most researchers to complete. In the last 10 years sequencing costs decreased drastically. The drastic decrease in sequencing price has made the methods outlined in this study practical for completing metagenomics analyses of bacterial communities.

Perhaps the most well-known microbiome study is the Human Microbiome Project (HMP). Microbial symbionts are present on every surface and in every nook, crevice, and cavity in the human body (Turnbaugh *et al,* 2007). Microbial cells outnumber mammalian cells ten-fold in an adult human being (Wilson, 2005). While some human microbes can be pathogenic and cause disease, others are vital for our survival. It is because of the vast majority of microbes and their many functional roles in humans that the HMP was completed.

**Land-based pollution sources and their impacts in Southeast Florida**

Land-based pollution sources are characterized as pollution into the marine environment from sewage outfalls, agricultural runoff, chemical spills, urban development, and oil spills (Windom, 1992). Land-based pollution is a considerable problem along the world's coastlines, with sewage discharge into coastal regions being one of the largest sources of marine pollution (Windom, 1992, and Meyer-Reil and Koster, 2000). Land-based runoff and nutrients enrich the coastal marine environment with pollutants, which can result in eutrophication (Lapointe *et al.*2004, Carrie Futch *et al,*2011, and Lapointe *et al,* 2015). Eutrophication can lead to conditions of hypoxia or anoxia and decreased biodiversity causing major shifts in the community structure of an ecosystem (Windom, 1992; Marcus, 2004, Lapointe *et al,* 2004). Community shifts result in an increase of phytoplankton, harmful algal blooms, and decrease zooplankton, bacterioplankton, and macro-organism diversity (Marcus, 2004, Paerl *et al,* 2006).

Microbes contribute the largest biomass to the marine environment, and are the main source for nutrient recycling (Paerl *et al,* 2006). Land-based point and non-point pollution sources can increase harmful marine pathogens, and cause a die-off of vital nutrient recycling bacteria (Sniezko, 1974).

Human development along coastlines is a significant contributor to marine pollution (Mallin *et al,* 2000, Lapointe *et al,* 2004, Carrie Futch *et al,* 2011, and Lapointe *et al,* 2015, ). The past decade has marked rapid increases in human population and coastal development, increasing sewage waste and construction debris, which directly impacts coastal marine environments (Mallin *et al,* 2000). One of the more serious threats of coastal development is the input of human pathogenic and enteric bacteria (McIntyre, 1990; Windom, 1992; Mallin *et al*, 2000). Pathogens enter the marine environment in the form of treated sewage, human excrement, and shedding of human skin cells (Mallin *et al*, 2000). The potential for harmful bacteria in coastal waters could continue to increase as urban development, tourism, and water-based activities gain popularity (Mallin *et al*, 2000). The introduction of enteric/pathogenic bacteria into the water can cause human illnesses such as: ear, respiratory, gastric, and skin infections (Mallin *et al*, 2000). Introduction of these pathogens to sensitive ecosystems are also reason for concern as recent research shows evidence that human pathogens are harmful to marine organisms (Banks *et al,* 2008 and Carrie Futch *et al.* 2011, Sutherland *et al,* 2011). This is observed with the human pathogen *Serratia marcescens,* which can cause serratiosis in the Caribbean elkhorn coral *Acropora palmata* (Sutherland *et al*, 2010). This can result in large bleaching events, and massive die-offs of the already threatened elkhorn coral (Sutherland *et al,* 2011).

Land-based pollution is a sizable problem in South Florida. With approximately 7 million inhabitants along Florida's southeastern coast, urbanization is intensifying (Finkl and Charlier, 2003, Banks *et al,* 2008, and Stamates *et al,* 2013). Two of the largest ports in the world are located on Florida's southeastern coast, and oil leakage, cargo spills, and flushing of ballast water all contribute to pollution along the coastline. Influxes of nutrients from agricultural practices in near-shore environments are believed to contribute to degradation of Florida's coral reefs (Finkl and Charlier, 2003, Banks *et al,* 2008,

Walker *et al,* 2008,2012). Large discharges of polluted freshwater are injected into estuarine environments, and can cause large harmful algal blooms, and loss of oyster reefs and seagrass beds (Finkl and Charlier, 2003). The presence of large sewage outfalls just offshore and along the coast of southeastern Florida adjacent to the coral reef ecosystems have caused harmful algal blooms and introduction of human enteric and fecal bacteria (Edmond *et al,* 1978; Finkl and Charlier, 2003). Land-based pollution has many unfavorable impacts on marine coastal environments. These impacts not only affect marine organisms and environments, but also humans utilizing these environments. It is important to consider these factors when studying the microbial communities present in an area largely impacted by urbanization and land-based pollution sources.

**Florida's Ports and Port Everglades Inlet**

Port Everglades Inlet is a man-made, dredged, deep-water port located along the southeastern coast of Florida (Stauble 1993, http://www.porteverglades.net; NOS, 2012). The entrance to the inlet is roughly 641 m long and 295 m wide with an average depth of 13 m (Stauble, 1993; NOS, 2011). The average tidal current speed is 0.7 knots, but can reach speeds up to 3 knots at flood tide, and 5 knots at ebb tide. Prominent winds that blow from the southeast and east that can travel at speeds of 17 knots or greater, causing very strong currents at the entrance to the inlet and can be hazardous to ships (NOS, 2011). Annual precipitation at the port is roughly 94 days per year, and 60% of this rainfall occurs in the summer months of May-October/November (NOS, 2011). The temperature average for the port is 32.2°C, with a maxima of 37.8°C in the summer, and -2.2°C in the winter (NOS, 2011). The seaport was officially established in 1927. In 1928 the barrier between the harbor and the ocean was demolished, opening the port for business. Port Everglades is a major source of income for Broward County, generating roughly $26 billion worth of revenue and business, and 201,000 jobs throughout the state (http://www.porteverglades.net). The main function of the port is for consumer purposes, but there is also a large amount of foreign trading (NOS, 2011). Port Everglades Inlet is mainly used for large vessel traffic with the major products handled in the port being petroleum, cars, cement, lumber, steel, glass, and other general cargo (NOS, 2011). In

2012 the total number of ships that docked in the port (excluding personal yachts and boats) exceeded 4000, making it one of the busiest cruise ship ports in the world (Carrie-Futch, 2011; NOS, 2011, and Stamates *et al,* 2013), and one of the most active cargo ports in the United States. Port Everglades is also the main seaport in southern Florida for petroleum products (http://www.porteverglades.net; NOS,2011;Carrie-Futch *et al,* 2011).

Port Everglades is constantly under human influence from cargo ships, cruise ships, naval ships, and recreational yachts and boats, as well as the confluence of runoff from over 1 million residents (Banks *et al,* 2008 and Carrie-Futch *et al,* 2011). The high volume of boat traffic and human influence, paired with the large amount of water discharged with the tides twice daily makes Port Everglades Inlet a possible point source of pollution to the offshore marine environments (Banks *et al,* 2008; Carrie-Futch, 2011; Stamates *et al,* 2013). A study by Carsey *et al.* (2007) in conjunction with NOAA reported that inlets in southern Florida provide substantial outflow of nutrients and microbes into the surrounding marine environment, and need to be monitored as a pollutant source. In particular, Port Everglades has been documented to have significantly high levels of nutrient and microbial contaminants discharged out of the port during ebb tide (Stamates *et al,* 2013). This emphasizes that Port Everglades inlet plays a possible role as a point source of pollution into the marine environment (Banks *et al,* 2008 and Stamates *et al,* 2013). The use of high-throughput sequencing coupled with water chemistry analysis will provide data regarding the types of bacteria present in the inlet and whether or not they can be potential threats to the marine and to humans who use the ocean surrounding the port for recreational purposes.

This study examines the surface water microbiome present in Port Everglades Inlet. Located directly offshore from the port is a sensitive coral reef tract (Rowher 2010), as well as multiple recreational beaches, fishing piers, and watersport areas (Banks *et al*, 2008 and Stamates *et al,* 2013). Culture-based methods are unable to fully capture the diversity of the microbiome of the port, therefore high-throughput sequencing will be used. High-throughput sequencing provides an in-depth view of the microbiome of the port. Results will provide information regarding the types of microbes being introduced into the coastal areas adjacent to the port, and how the microbial composition will change

throughout a year. These results can be used in conjunction with the county or other organizations as a baseline measurement for comparative studies.

## Section II: Objectives and Hypotheses

In this study the microbial communities present in Port Everglades Inlet will be analyzed using high-throughput DNA sequencing to complete a microbiomic study via 16S amplicon library analysis. The purpose of this study is to determine the surface water microbial population of Port Everglades Inlet (PEI) based on season (wet vs. dry), water chemistry and site type, from water samples taken weekly over a one year timespan. The analysis was done using high-throughput sequencing techniques on Illumina's MiSeq platform to eliminate the need for culture-based techniques. Sequencing data was analyzed with the bioinformatics programs QIIME and the R statistical software package Phyloseq.

The first hypothesis for this study predicts that during the typical wet season (May-September) there will be an increased abundance of bacterial species present in the port's water. The second hypothesis predicts that changes in water chemistry will correlate with changes in abundance of certain microbial genera. Species such *Prochlorococcus* are expected to be higher in abundance during low nutrient conditions (Flombaum *et al,* 2013; Johnson *et al,* 2006), while species such as *Synechococcus* are expected to be present in higher abundance during high nutrient conditions (Partensky *et al,* 1999). The third hypothesis predicts that harmful pathogens to both humans and marine life such as *Enterococcus, Streptococcus, Vibrios,* and *S. marcescens* will be present, and that these organisms will be present in a higher abundance during the wet season.

**Section III: Methods**

**Water sample collection and filtration:**

Surface seawater samples were collected by kayak in Port Everglades Inlet located in Broward County, FL at the beginning of low tide. Samples were collected on a weekly basis for one year (June 2013-June 2014). Three samples were collected from the surface water at two different sites within the inlet in 1 liter (L) glass Pyrex bottles. Water samples were collected for DNA sequencing and water chemistry analysis . Water temperature was measured in-situ at time of sampling with a glass thermometer. Salinity measurements were taken immediately upon returning to the laboratory using a refractometer. The 1L samples were then filtered using a vacuum filtration system and Pall GN Metricel® grid 47 mm, 0.45µm filters. The vacuum filtration system used Pall funnel filters and a rubber stopper to create an airtight seal. After filtration the filters were folded and placed into 2.8 mm ceramic bead tubes obtained from the MO BIO (Carlsbad, CA) using sterile forceps. The filters were folded and stored in -80°C freezer until DNA extraction. A total of 82 samples were prepared for sequencing.

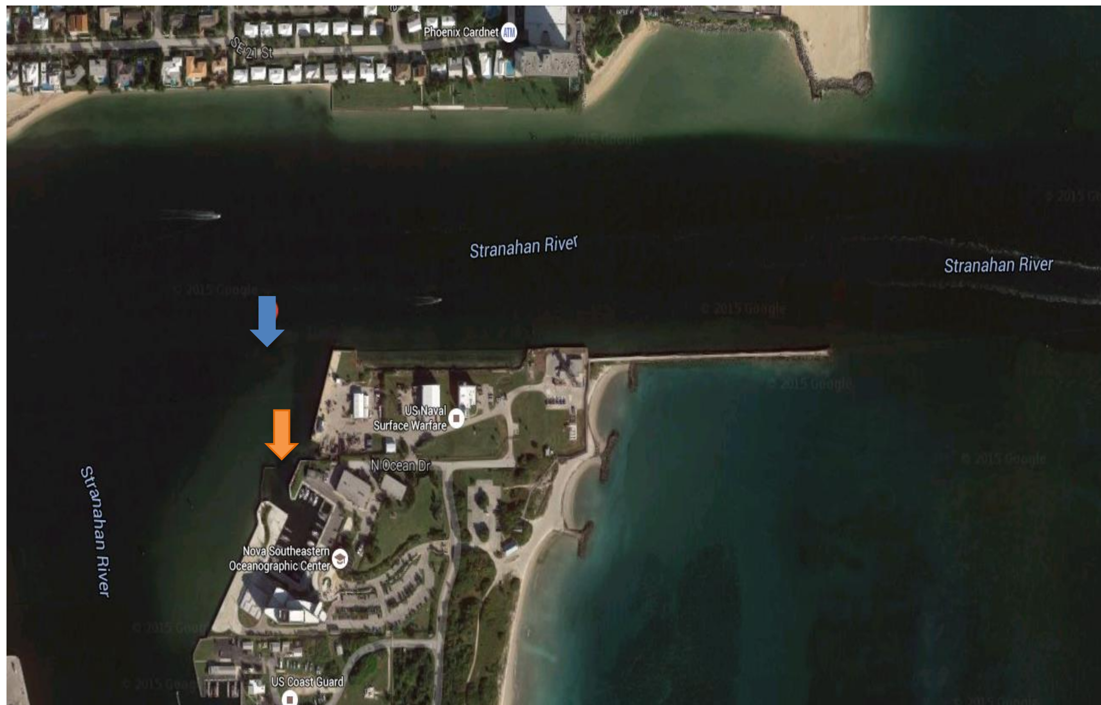


**Figure 1**: Port Everglades Inlet sampling locations. The blue arrow is Marker 9 (M9) and the orange arrow is the boast basin (BB). Image obtained from Google Maps (https://www.google.com/maps/place/26%C2%B005'33.6%22N+80%C2%B006'44.0%22W/@26.0920736,-80.1086591,1023m/data=!3m1!1e3!4m2!3m1!1s0x0:0x0)

| Site Type | Latitude | Longitude |
|---|---|---|
| Boat Basin (BB) | 26.091510 | -80.112098 |
| Marker 9 (M9) | 26.092675 | -80.112216 |

**Table 1**: Latitude and longitude of sites where water samples were collected

**DNA extraction:**

Microbial genomic DNA was extracted using MO BIO's PowerLyzer™ PowerSoil® kit, generally following Earth Microbiome (earthmicrobiome.org) standards. After extraction a 1% agarose gel was run to ensure that the DNA extraction was

successful. After gel verification the purity of the DNA was checked using the NanoDrop 1000 (Thermo Scientific), and DNA concentration was confirmed using the Qubit 2.0 (Life Technologies).

**Ion Chromatography (For Water Chemistry Analysis):**

Surface seawater samples were collected in 50 mL conical vials at each site in the inlet and diluted 1000X before injection into the Ion Chromatography machine. Ion chromatography (IC) is an analytical method used to determine the presence and concentration of chemical ions in solution, including phosphate, chloride, sulfate, nitrate and fluoride ions for this study. IC methods were completed by Dr. Song Gao, Associate Professor of Chemistry at NSU's Farquhar College of Arts and Sciences (until July 1, 2015) and then Halmos College of Natural Sciences and Oceanography (after July 1, 2015). The IC analysis was performed using a Thermo Scientific Dionex ICS-1600 (Bannockburn, IL).

**Illumina MiSeq High-throughput sequencing sample preparation:**

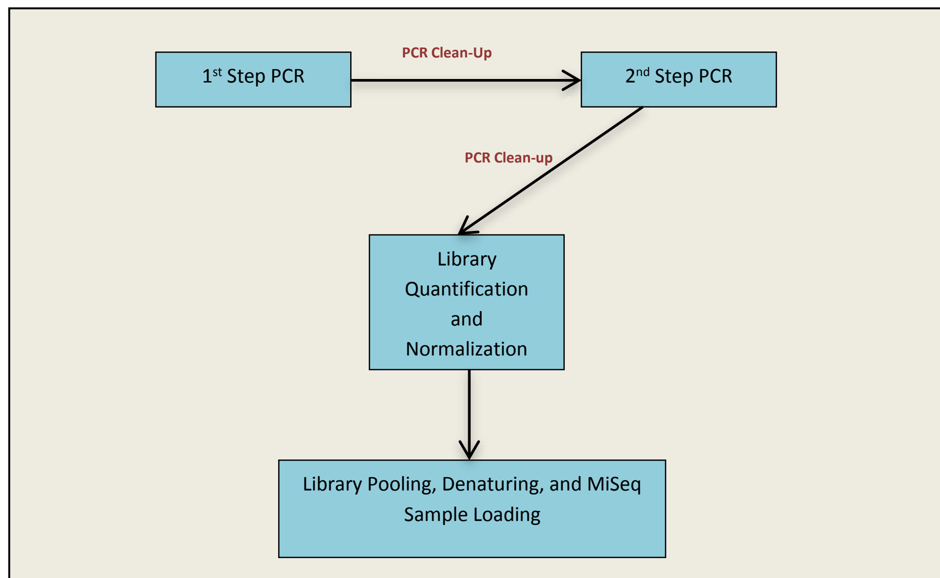


**Figure 2**: Workflow of sample preparation for 16s metagenomic sequencing

Preparation of seawater samples for sequencing followed Illumina's 16S Metagenomic Sequencing Library Preparation guide (Illumina, 2013). The workflow consisted of a two-step PCR, library quantification and normalization, and pooling of the libraries (Fig. 2).The amplicon primers used for the first step PCR were the universal primers, MIDf-515F and 806rc (Caporaso *et al,* 2011) which amplified the V4 region of the 16S gene. Attached to the locus-specific primers were forward and reverse overhang sequence adapters determined by Illumina. The overhang sequences were added so that the Illumina Nextera XT indices could attach to the locus-specific primers. The genomic DNA concentration for each sample was determined using the Qubit for normalization. Samples that had a concentration higher than 5 ng/μL were diluted to that value before running the PCRs. The PCR conditions for the first-step amplicon PCR were as followed:

The initial denaturation step was 95°C for 3 minutes. The initial denaturation step was followed by denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, and finally, extension at 72°C for 30 seconds. These steps were repeated for 25 cycles. The final step was extension at 72°C for 5 minutes and the reactions were held at 4°C indefinitely. The PCR products were cleaned-up using AMPure XP beads (Agencourt) and a magnetic plate as outlined in the 16S metagenomic library prep guide (Illumina, 2013). The second-step PCR or index PCR attached the Nextera XT dual indices and the Illumina sequencing adaptors using the Nextera XT indexing kit (Illumina). The PCR conditions were as followed: The initial denaturation step was at 95°C for 3 minutes. Following the initial denaturation step a second denaturation occurred at 95°C for 30 seconds. Then annealing occurred at 55°C for 30 seconds, followed by extension at 72°C for 30 seconds. These steps were repeated for 8 cycles. The last step was a final extension at 72°C for 5 minutes, and then reactions were held at 4°C indefinitely. The PCR products were cleaned-up using Ampure XP beads and a magnetic plate following the procedure outlined in Illumina's 16S metagenomic library prep guide (Illumina, 2013).

After the completion of the two PCR reactions, library quantification, normalization, and pooling were completed. To calculate the DNA concentration in nM, average library size was determined using the 2200 TapeStation (Agilent Technologies) and DNA concentration was determined using the Qubit 2.0. High Sensitivity D1000 ScreenTapes were used for the TapeStation. Samples were prepared for the TapeStation by following the protocol outlined in the Agilent High Sensitivity D1000 ScreenTape System Quick Guide (Agilent Technologies, 2013). Library quantification was determined by using the following calculation:

**(concentration in ng/µL)/(660 g/mol x avg. library size) x $10^6$ = concentration in nM**

The libraries were diluted with 10mM Tris pH 8.5 to a final concentration of 4 nM for normalization. A total of 5 µL of DNA per library (sample) was mixed together for multiplexing of libraries.

Library denaturation and dilution was completed as outlined in the 16S metagenomics library preparation guide (Illumina, 2013). The final pooled DNA library

was diluted down to 4 pM and a 50% spike in of 12.5 pM PhiX was added to the pooled library. The 50:50 DNA library and PhiX control mix was loaded into a MiSeq V3 600 (2 x 300) cycle kit and run on the MiSeq using targeted resequencing. The 16S metagenomics workflow was used and adaptor and primer trimming occurred previous to file generation. Upon the completion of sequencing, 2 FASTQ files, a forward and a reverse read, were generated for each of the 82 samples and were used for downstream analysis.

**Sequencing Analysis: QIIME Analysis:**

Raw sequences were analyzed using Quantitative Insights into Microbial Ecology (QIIME) version 1.8.0 (Caporaso *et al.* 2010). The forward and reverse paired-end reads were joined and converted to FASTA files using the program PANDAseq version 2.8.1(Masalla *et al,* 2012), with a confidence level of 90%. A mapping file was made containing all sample names and metadata for each sample. The mapping file consists of all samples and their associated metadata necessary to complete downstream analysis (http://qiime.org/documentation/file_formats.html#mapping-file-overview). The mapping file contains sample ID, linker primer sequence, and multiple description columns containing metadata (salinity, temperature, season, etc). The assembled paired-end reads were merged into one large file to be used in downstream analysis. Chimera checking was done using USEARCH 6.1 (Edgar, 2010), and chimeric sequences were removed from the dataset. Operational taxonomic units (OTUs) were picked using the cd-hit method (Li and Godzik, 2006), which works by using a "longest-sequence-first list removal algorithm" in order to group similar sequences. Pre-filtering steps were completed so that identical sequences, or those sequences that are sub-sequences of other sequences, were collapsed previous to OTU picking. During OTU picking, sequences with 97% sequence similarity were grouped into the same OTU. After OTUs were established a representative sequence was chosen from each OTU by using the default method in QIIME, which uses the first sequence listed in each OTU. Once the representative sequence was chosen for each OTU, the sequences were aligned to a database. The sequences were aligned using PyNAST (Caporaso *et al,* 2009) in QIIME. PyNAST uses a "gold" pre-aligned template made from the greengenes 13_8 reference

database (DeSantis *et al.* 2006; McDonald *et al,*2012). Sequences with less than 75% sequence identity with a minimum length of 150 bp were discarded and not used in downstream analysis. After alignment, sequences were assigned taxonomy using the RDP classifier with the greengenes 16S rRNA database. After taxonomy was assigned, an OTU table and a phylogenetic tree were generated. An OTU table was generated in the .biom file format, and the phylogenetic tree file was generated and saved in Newick format. These file types were used in downstream analysis in R statistical software(R Core Team, 2013).

**Sequencing/Statistical Analysis: QIIME and R**

*Importing, Filtering, and Normalization*

The package phyloseq (McMurdie and Holmes, 2013) was used for downstream statistical analysis requiring the input of the OTU table and phylogenetic tree files generated in QIIME. The .biom file and the .tre file were imported into R and merged into one data matrix using the import_biom command. The mapping file was imported into R and merged with the OTU table and phylogenetic tree data matrix using the merge_phyloseq command. Rank names were checked to make sure that the data matrix contained all taxonomic ranks using the rank_names command. The parse_taxonomy_greengenes command was used to generate taxonomy and produce proper taxonomic rank names. An extra rank name was added to the data matrix and was removed using the tax_table command. A new phyloseq object was generated to include only taxa with a relative abundance of greater than 1%, to exclude rare microbiota for beta diversity analysis. This generated a phyloseq object containing the core microbiome of surface water samples in the Port Everglades Inlet. Data was normalized to the smallest library size of all the samples, which was 5666 reads, using the transform_sample_counts command. After normalization, rank abundance curves were generated for the top 100 OTUs.

*Alpha Diversity QIIME*

*Rarefaction Curves:*

Rarefaction curves were generated in QIIME using the make_rarefaction_plots.py command. Rarefaction curves are used to compare diversity between samples using plots to determine if existing diversity was fully captured. The curves plot the number of sequences per sample by a specified diversity index. A steep curve that does not reach a plateau indicates that sampling depth was insufficient to resolve full bacterial diversity. A steep curve reaching a plateau indicates that sampling depth was sufficient to capture full bacterial diversity and no further sampling or re-sequencing is needed.

*Alpha Diversity Phyloseq*

Alpha diversity was measured by first removing all OTUs with counts of zero. The data was then rarefied to 5666 reads for all samples with replacement 100 times using the rarefy_even_depth command in order to analyze all samples at an even sequencing depth. Richness and evenness estimates were determined using the plyr package and the apply command (Wickham, 2011). Shannon, Simpson, Observed, and Chao1 alpha diversity estimates were generated and plotted. Community composition barplots were created using the top 500 OTUs for phylum and genus and the top 100 OTUs for class, order, and family using the ggplots (Wickham, 2009) package in R. A non-parametric kruskal-wallis test was used to complete pairwise comparisons to determine significant differences in the alpha diversity between month, location, and the interaction term month and location.

*Beta Diversity*

Beta diversity analysis was conducted by calculating both Bray-Curtis distance and weighted Unifrac distance in the phyloseq package. Bray- Curtis distance quantifies

compositional dissimilarities between site, month, and season, based on the number of counts for those criteria (Faith *et al,* 1987). Weighted Unifrac distance is a quantitative measure that determines changes in abundance (Lozupone *et al,* 2007). After calculating both Bray-Curtis dissimilarity and weighted Unifrac values, ordination plots were generated. PCoA plots and NMDS plots were generated for location, month, and season. The non-parametric Adonis test was used to complete pairwise comparisons of the samples for month, location, and season. A p-value less than 0.05 is considered significant and this statistic tells you how likely it is your results occurred by chance. The $R^2$ value is used in Adonis testing and provides a value that indicates how much of the variation in the distances is explained by the grouping being tested.

*Pathogen Detection*

Potentially pathogenic bacteria were detected by filtering out the orders known to contain pathogens of interest using the subset_taxa command in phyloseq. The filtered orders were pruned to contain only the top 50 OTUs from the subset of data in the previous step. The abundance, genus, and month were plotted using the plot_bar command in phyloseq.

**Regression Analysis with Environmental Metadata**

A series of Multiple least Squares Regression's were used to assess possible relationships between each bacterial taxa and the environmental variables taken as part of the study. A backward Selection method was used with both entry and model retention set at alpha=0.10. All regression analyses were carried out using SAS Statistical Software (SAS Institute).

## Section IV: Results

### MiSeq Output

A total of 151 samples were collected weekly from Port Everglades Inlet from July 2013-June 2014. A total of 42 samples from the marker 9 site and a total of 40 samples from the boat basin site were used for DNA sequencing resulting in an overall total of 82 samples that were sequenced. The V4 hypervariable region of the 16S rRNA gene was sequenced due to evidence that it generates the most accurate taxonomic results when using Illumina technology (Shah *et al,* 2010, He *et al,* 2013 and Fadrosh *et al,* 2014). The total number of raw DNA sequences generated from MiSeq sequencing was 1,435,072 with Q scores greater than 30. The average number of reads per sample was 17,287. The minimum number of reads per sample was 5,666 and the maximum number of reads per sample was 80,122. Sequences were checked for chimeras using USEARCH 6.1. Chimeric sequences are sequences composed of DNA from two or more parent sequences. The sequences, if not removed, can be classified as novel sequences, when they are truely PCR artifacts. A total of 25,020 sequences were removed after chimera checking. Filtering and OTU picking was done with the remaining 1,410,052 sequences. The CD-HIT method was used for OTU picking because it is recommended for datasets with larger than 100,000 sequences. Pre-filtering was completed to remove identical sequences and subsequences. After filtering, a total of 395,009 unique sequences were left with an average amplicon size of roughly 250 basepairs. OTU picking was completed with a cluster cutoff of 97% pairwise identity. A total of 16,384 OTUs were generated from 395,009 sequences**.** Taxonomy was assigned using the Greengenes 13_8 database and sequences were aligned using PyNAST. A total of 2,555 sequences did not align to the reference database and were removed from the analysis. The remaining sequences were for all further downstream analyses.

## Alpha Diversity

***Rarefaction Curves:*** Rarefaction curves were generated in QIIME and phyloseq to determine if sampling depth was sufficient to fully capture and characterize diversity of the water samples by plotting alpha diversity against simulated sequencing effort. Rarefaction curves were generated for both Chao1 richness estimator and Shannon diversity index in QIIME and were rarefied to 15000 reads (Figs. 3 and 4). Curves for

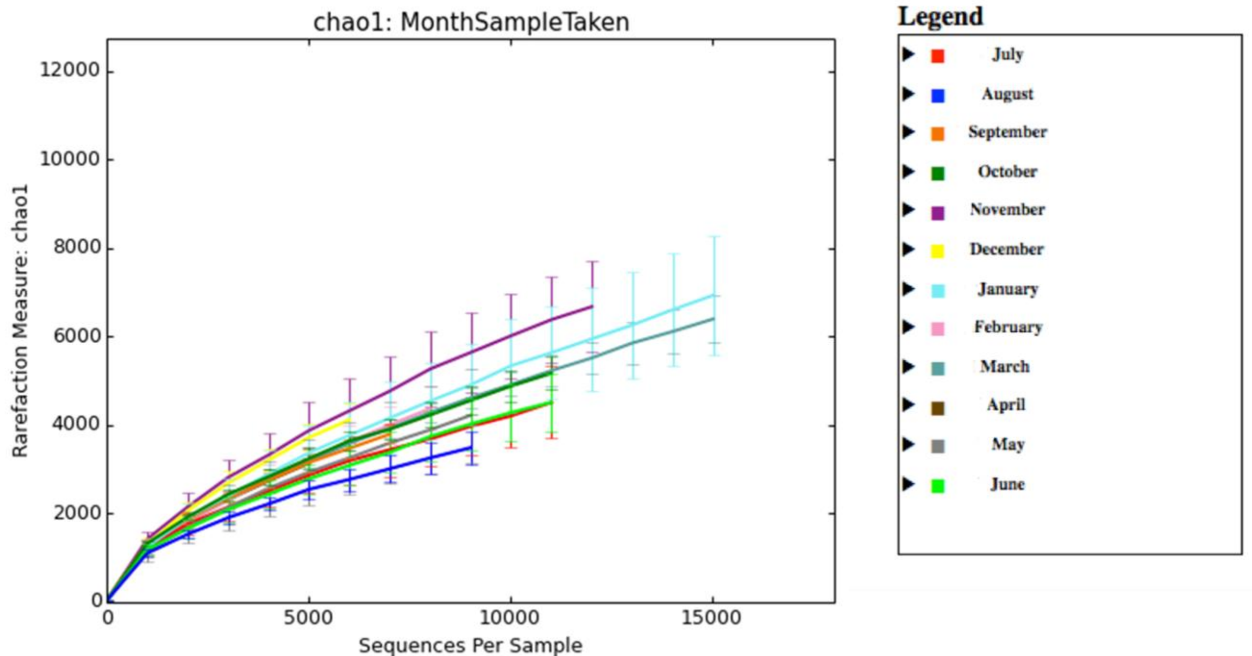


**Figure 3:** Rarefaction curves of month for Chao1 diversity indices. Curves generated in QIIME 1.8.0.

Observed species richness and Shannon diversity were generated in phyloseq without rarefaction. The non-rarefied curve and rarefied curves both show similar results when looking at diversity indices. The Chao1 curve, which accounts for both species richness and abundance, shows a steep incline with gradual flattening of the curve. The curves do not reach a full plateau indicating that full bacterial diversity may not be represented. The observed species curve generated in phyloseq also accounts for both species richness and

diversity and shows a similar distribution (Fig. 5). The Shannon diversity curves, which account only for species richness, display steep curves with complete plateaus indicating the likeliness that full bacterial diversity was captured (Figs. 3 and 5). The samples that did not reach diversity saturation were included in further downstream analyses.

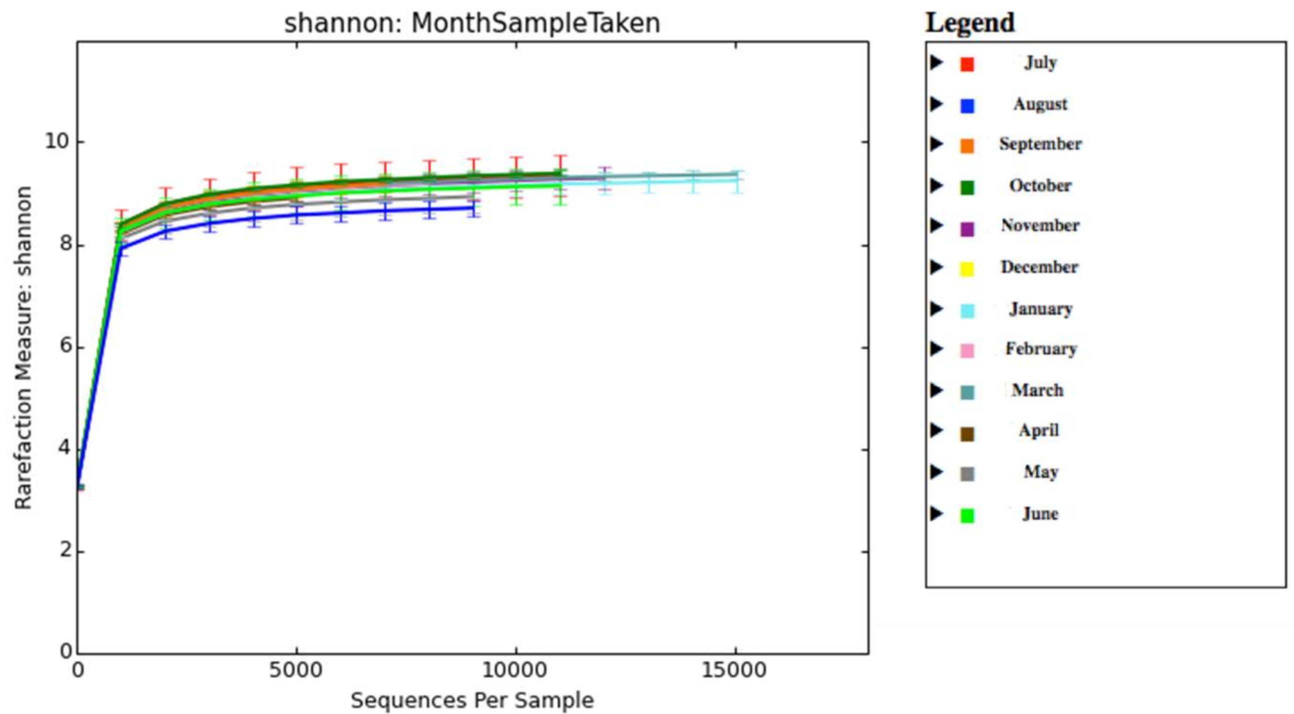


**Figure 4:** Rarefaction curves of month for Shannon diversity indices. Curves generated in QIIME 1.8.0.

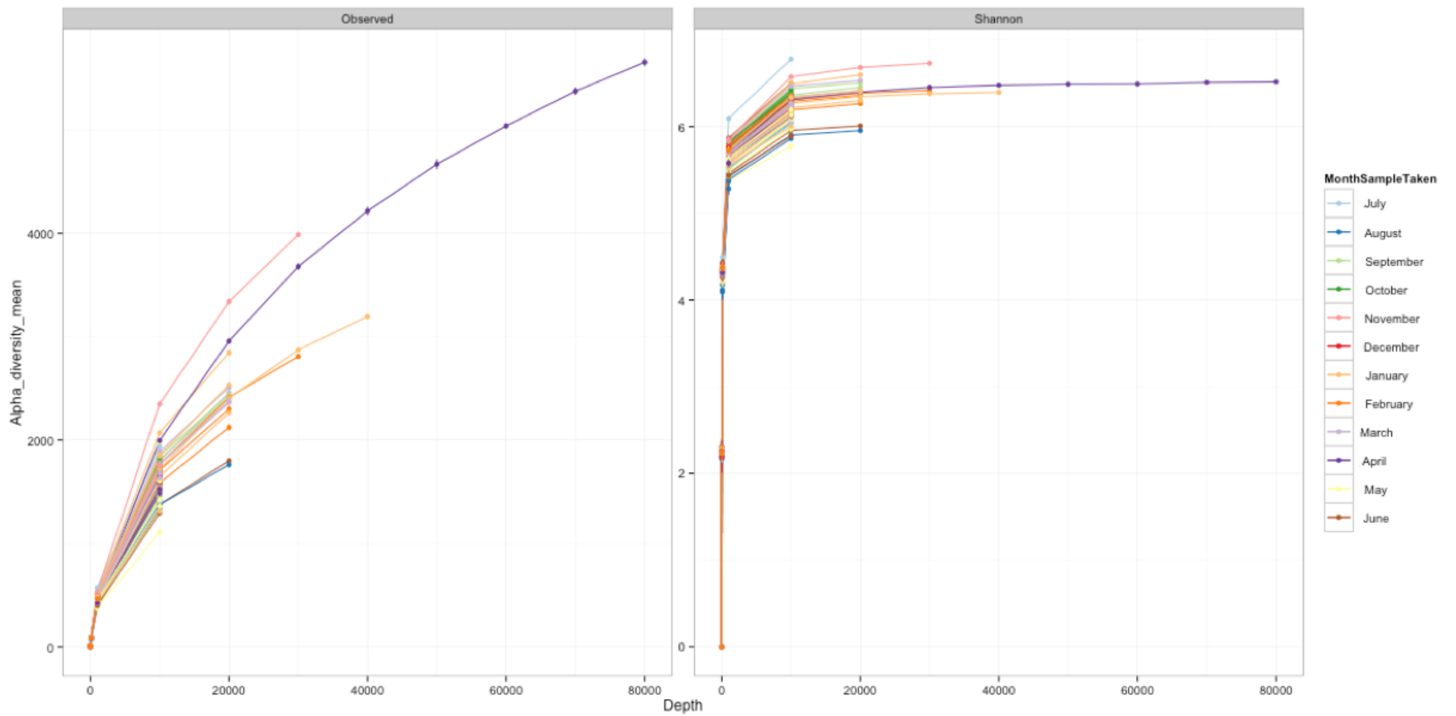


**Figure 5:** Rarefaction curves of month for Observed Species and Shannon diversity indices. Curves generated in Phyloseq

### ***Rank Abundance Curve***

Rank abundance curves depict relative abundance and species diversity within a community by plotting relative abundance of species against their rank in abundance

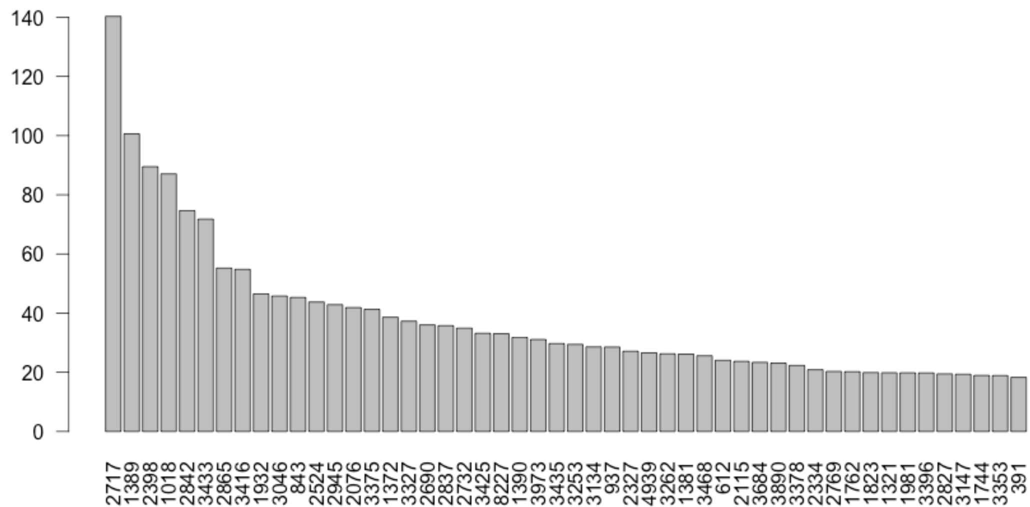


**Figure 6**: Rank abundance plot of top 500 OTUs in dataset. The top 10 OTUs are Rhodobacteraceae (2717), Cryomorphaceae (1389), Synechococcaceae (2398), Unclassified bacteria (1018), Stramenophiles (2842), Rhodobacteraceae (3433), Synechococcaceae (2865), Halomonadeceae (3416), Halomonadeceae (1932), Alphaproteobacteria (3046).

(Fig. 6). Rank abundance curves provide information at a glance regarding community composition. A high diversity environment will have a gradual curve with a few high abundance species and a long tail indicating the presence of singletons. A lower diversity curve will have a steep slope with a shorter tail comprised of singletons. The rank abundance curve displays the top 50 OTUs in the Port Everglades Inlet samples. The slope of the curve is relatively gradual indicating that the diversity of the water samples is comparably high. The most abundant OTU (2717) is classified as being the family Rhodobacteraceae in the phylum Proteobacteria. The second most abundant OTU (1389) is classified as being the family Cryomorphaceae in the phylum Bacteroidetes.

***Alpha Diversity Indices***

Port Everglades Inlet surface water samples were analyzed by high-throughput MiSeq16s DNA sequencing to determine total microbial composition. Quality filtering of sequences was completed and community richness and diversity measures were calculated for all samples. Alpha diversity is species diversity or richness within a sample, community, or habitat.

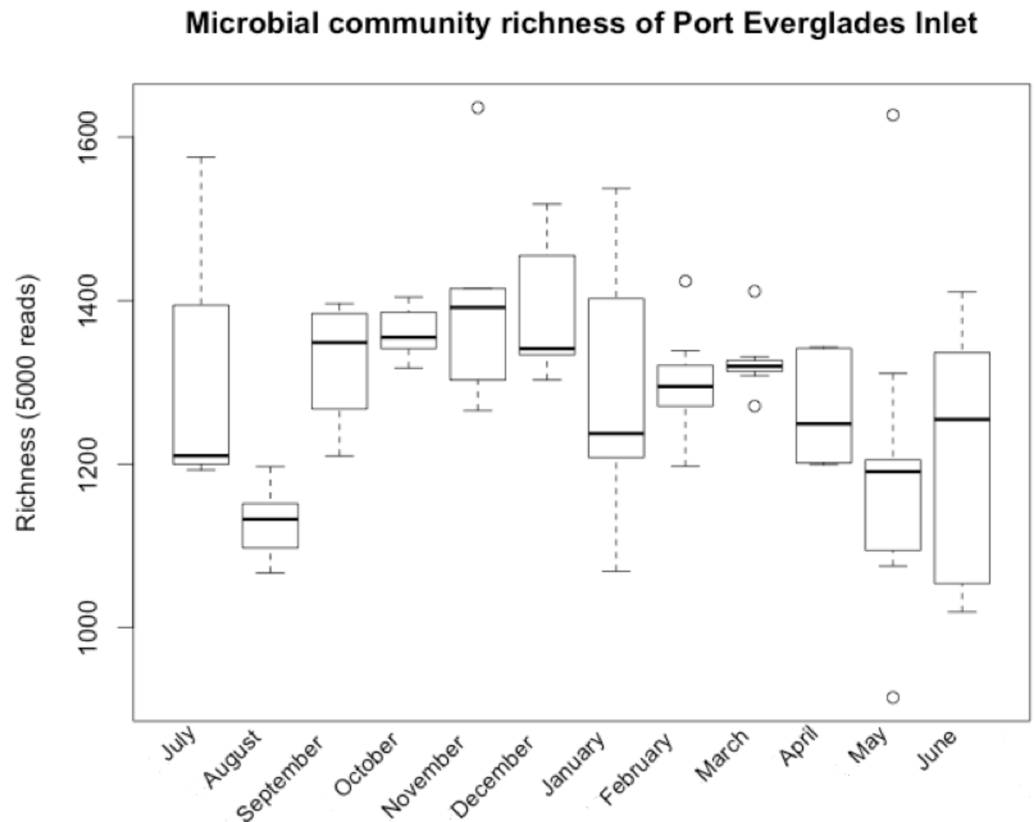


**Figure 7:** Community richness boxplot by month. Data rarefied to 500 reads per sample

To compare alpha diversity metrics non-parametric (Monte Carlo) two-sample t-tests were run. No significant differences (P>0.05) were observed between alpha diversity metrics for the t-tests (Chao1,Shannon, Simpson, Observed; Data in supplementary index). The month of August showed the lowest amount of species richness, while the months of November and December showed the highest amount of species richness (Figs. 7 and 8).

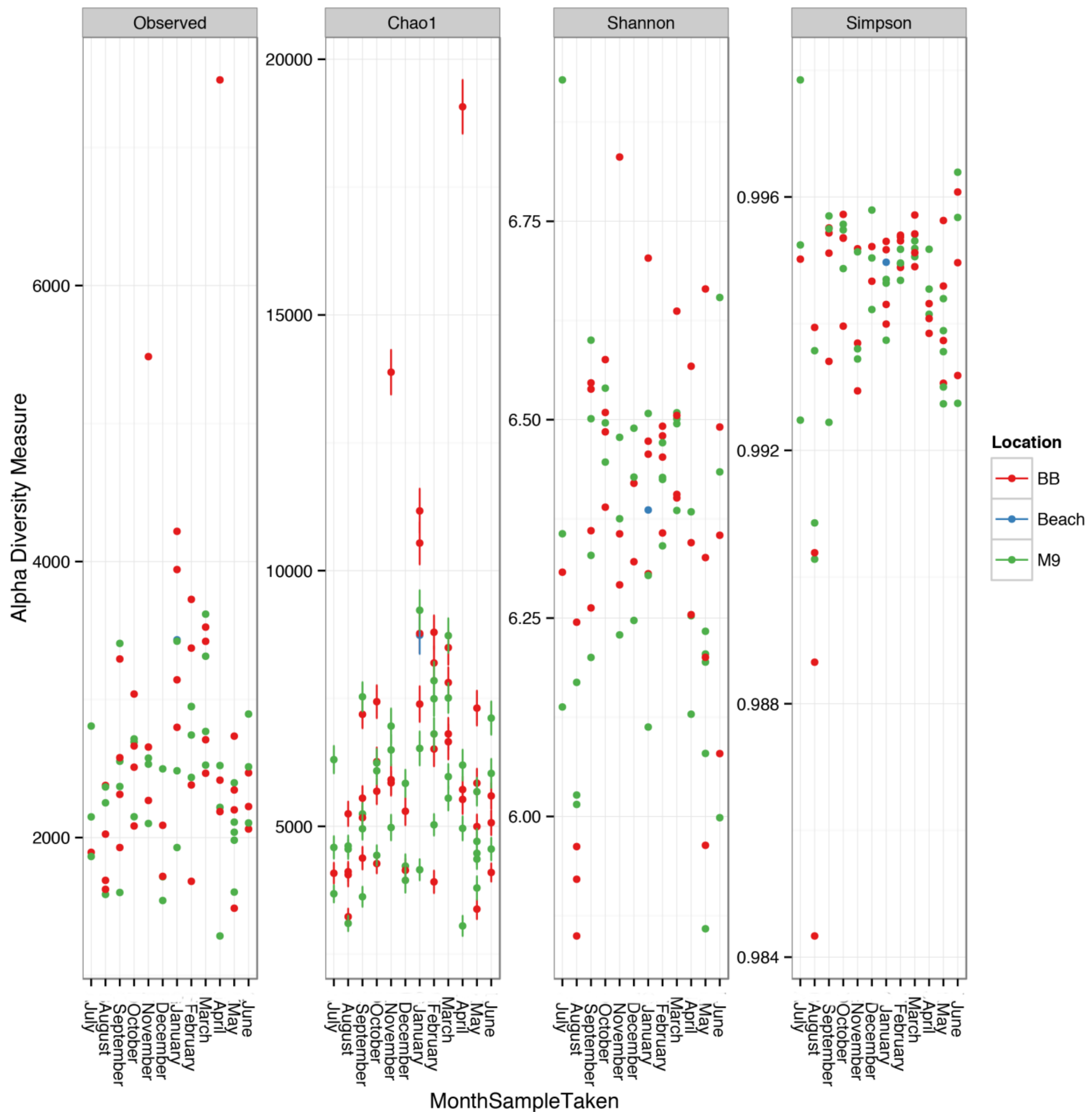


**Figure 8:** Alpha diversity index plots. Points are colored by sample location.

***Relative Abundance of Microbes in Port Everglades Inlet***

Relative microbial abundance of Port Everglades Inlet water samples are depicted using stacked bar charts and a frequency vs. relative abundance scatter plot. The scatter plot illustrates the 6 most abundant bacterial phyla found in all of the sequenced samples along with the frequency in which they are observed in the samples. The top 7 most abundant phyla are Actinobacteria, Cyanobacteria, Proteobacteria, Bacteroidetes, Euryarchaeota. Firmicutes, and Verrucomicrobia (Fig. 9). The majority of the most

common phyla are in the Bacterial domain, except for Euryarchaeota, which falls in the Archaeal domain.

Stacked taxa summary bar charts depict the relative abundance of microbes at different taxonomic levels for the top 1000 OTUs at the phylum level and the top 500 OTUs at the class and genus level for organisms present in greater than 1% in all samples. A Kruskal-Wallis test was completed to see if there were significant differences in community composition between months, locations, or seasons. The Kruskal-Wallis test showed significance when comparing the month of August with the months of December, October, and November as well as between seasons. No significant differences are observed by location. Taxonomic classification of the top 1000 OTUs at the phylum level comprises 96% of the full dataset and shows that Proteobacteria dominate the microbial species composition in all months except for August, September, and October, with the percent composition ranging from 34-51% (Fig. 10). Cyanobacteria dominate the species composition in August, September, and October comprising 38-45% of the samples in those months (Fig. 10). Bacteroidetes is largely present in water samples for all months ranging from 16-34%

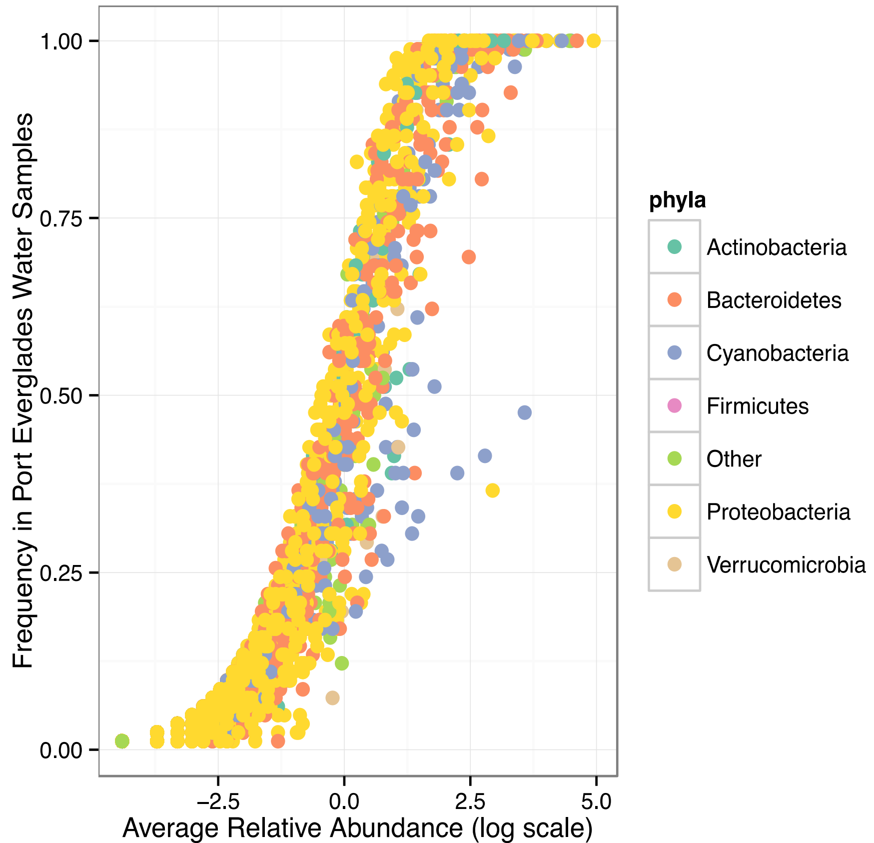


**Figure 9**: Average relative abundance vs. frequency plot for top 6 most abundance bacterial phyla in dataset

composition, with fluctuations in species composition corresponding with fluctuations in Cyanobacteria composition (Fig. 10). Actinobacteria is present in all water samples at a lower abundance ranging from 2-4% in the samples (Fig 10). Seasonal trends in species composition can also be seen with Proteobacteria and Bacteroidetes being most abundant during the dry season and Cyanobacteria being most abundant during the wet season (Fig 10).

The top 500 OTUs were used to generate species composition bar plots at the class and genus taxonomic levels. The top 500 OTUs represent 90% of the total dataset for organisms present in greater than 1% in all samples. The most abundant group at the class level is Flavobacteriia followed by Alphaproteobacteria (Fig 11). Seasonal trends are observed in the family Synechococcophycideae in which abundance doubles in the wet season months.

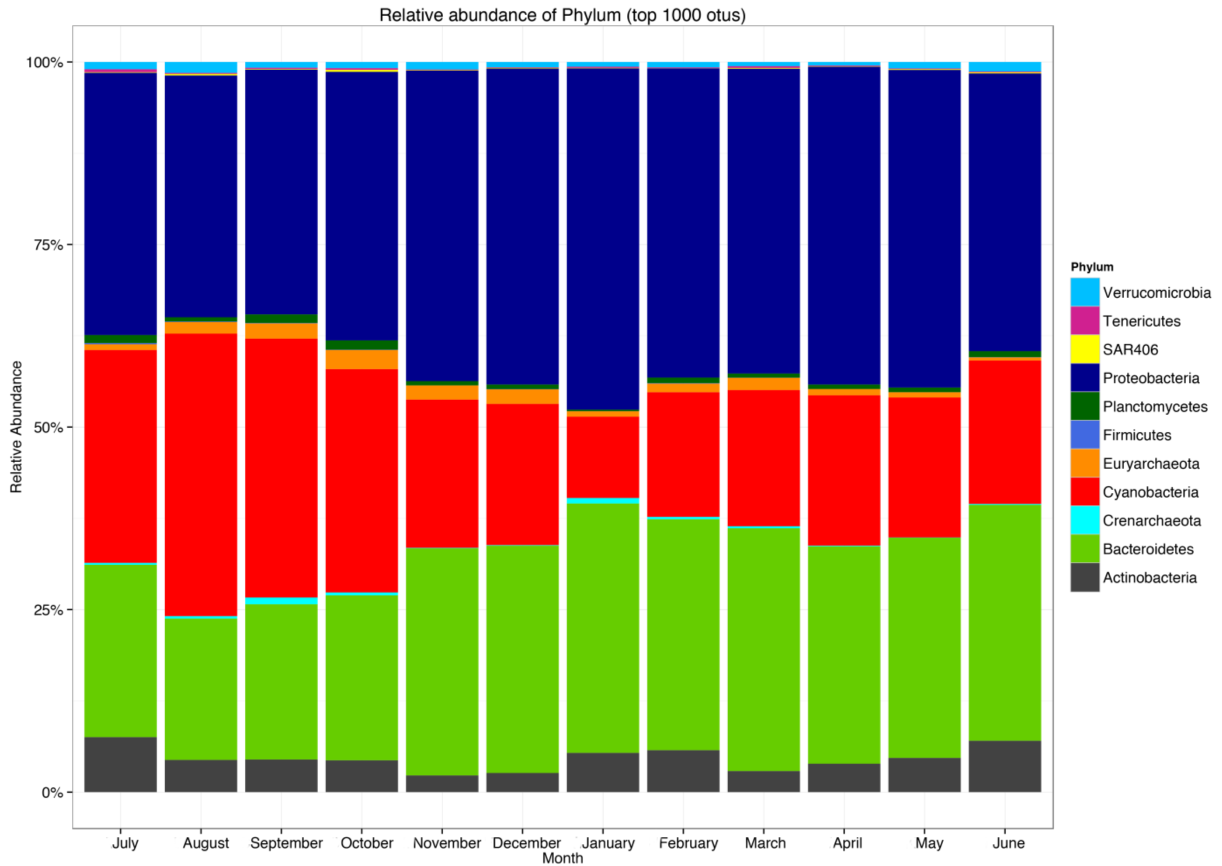


**Figure 10:** Taxa summary of relative abundance at phylum taxonomic level

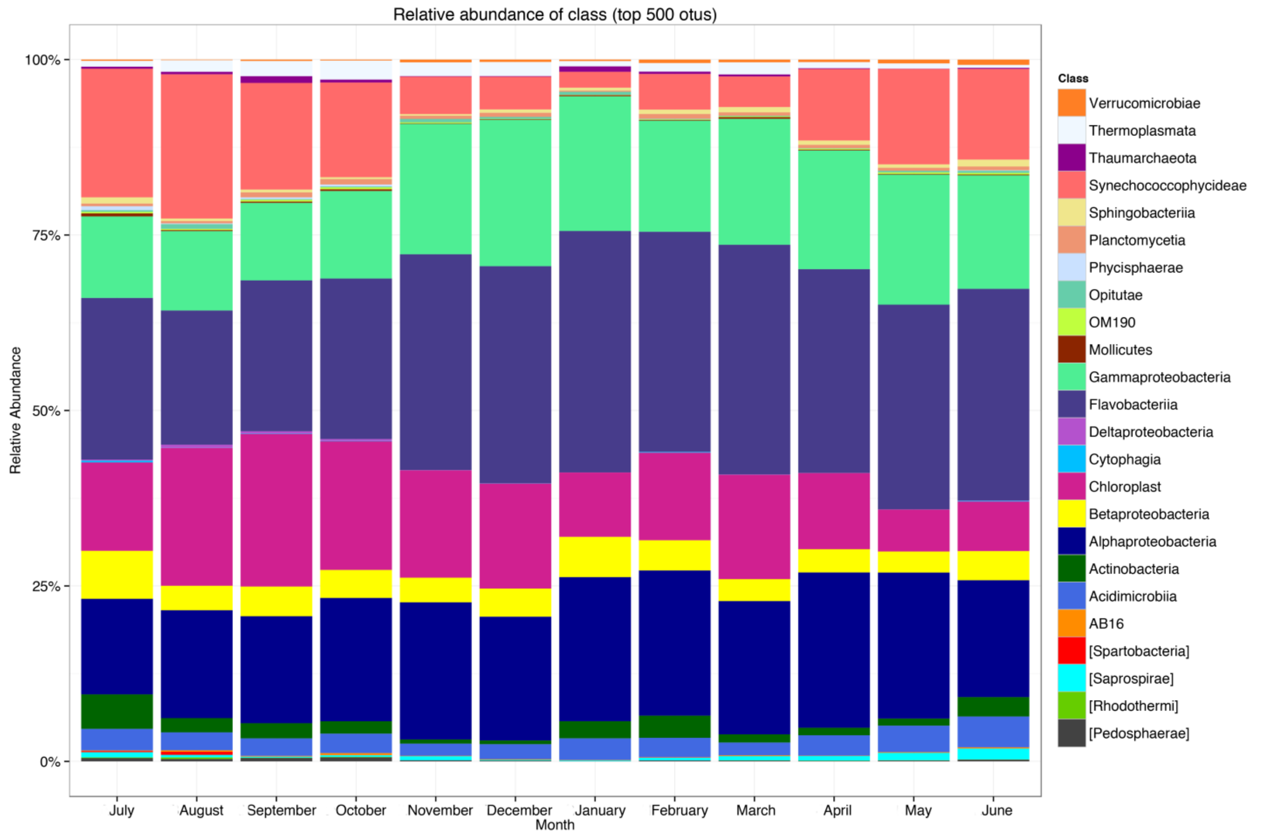


**Figure 11**: Taxa summary of relative abundance at class taxonomic level

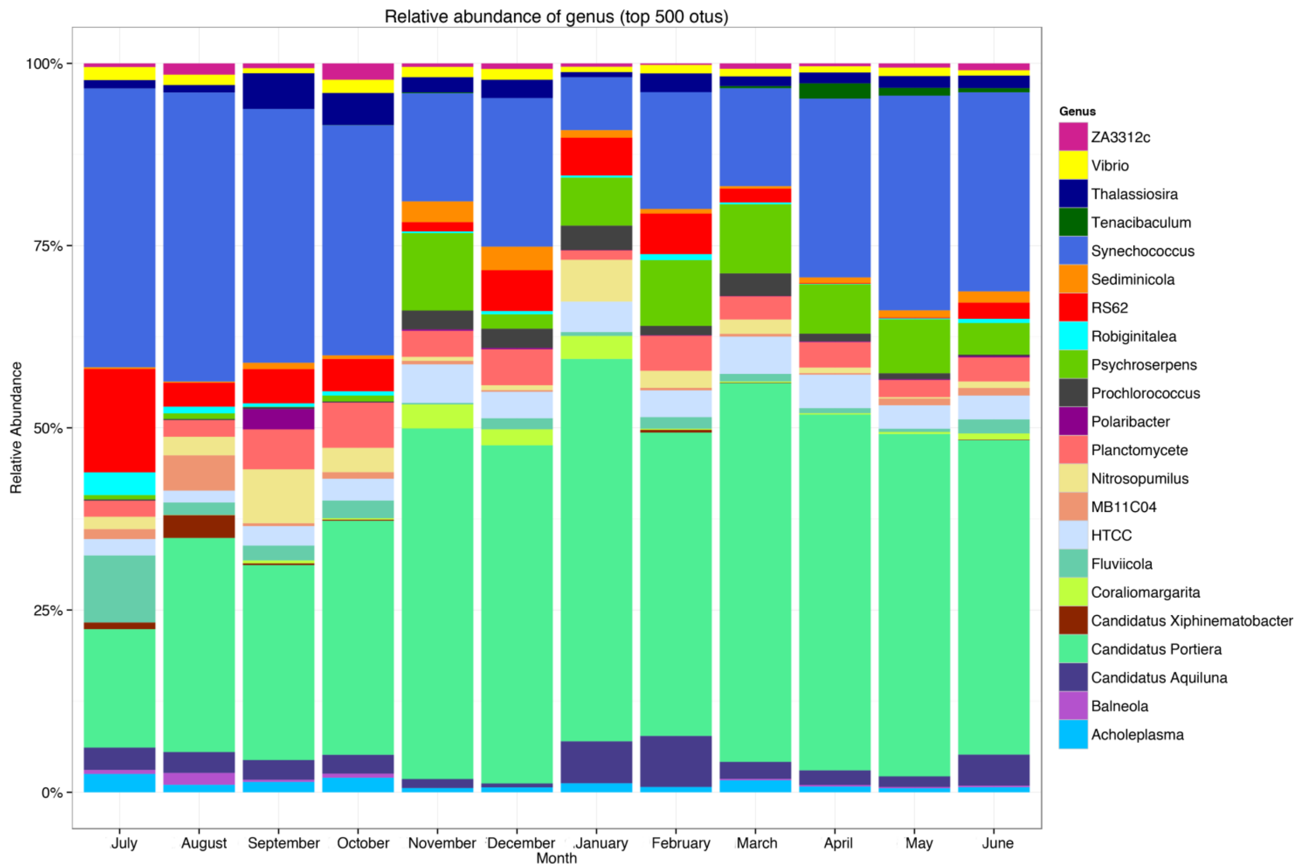


**Figure 12**: Taxa summary of relative abundance at genus taxonomic level

The most drastic shifts in community composition are observed at the genus taxonomic level (Fig 12). *Candidatus Portiera* and *Synechococcus* largely dominate the microbial abundance at this level. *Synechococcus* shifts in abundance are seen with changes in season, with the wet season having the highest abundance. The presence/absence of different taxa in different months is evident at the genus taxonomic level (Fig. 12). The taxon *Psychroserpens* is present in samples from November of 2013 to June of 2014, but is not present in samples taken from July to October of 2013. The presence of *Coraliomargarita* is seen in higher abundance in the months of November, December, and January but is lacking, or very scarce in the other months of the year (Fig. 12). The presence of *Candidatus Xiphinematobacter* is only seen in the months of July and August, while *Polaribacter* is only seen in September (Fig. 12).

## Beta Diversity

Beta diversity is used as a measure to compare samples to each other by calculating distances between pairs of samples. Beta diversity for Port Everglades Inlet surface water samples was determined by calculating both Bray-curtis dissimilarity and weighted Unifrac distance. These data are portrayed using Non-Metric Multidimensional Scaling (NMDS) ordination plots, which compare groups of samples using phylogenetic or count-based distance measures. Adonis tests were run on samples to analyze the strength of significance that a specific group has in determining variations in distances between samples. Adonis test are often used to analyze differences in composition and relative abundance of different species that comprise different samples, sites, etc. Results were considered significant is less than the value for Alpha at 0.05.

### *Location*

Water samples were taken from two locations in Port Everglades Inlet. Changes in beta diversity were analyzed by calculating Bray-Curtis dissimilarity and weighted Unifrac distance (Figs. 13 and 14). The results of the both the weighted Unifrac and Bray-Curtis dissimilarity show no distinct clustering of samples, meaning that there are no differences in beta diversity by location. To analyze this statistically an Adonis tests were run for both Bray-Curtis and weighted Unifrac distance and results concluded that there were no significant differences between beta diversity for location ($P > 0.05$, $R^2=0.0137$ (Bray-Curtis); $P > 0.05$, $R^2=0.0146$(Unifrac)).

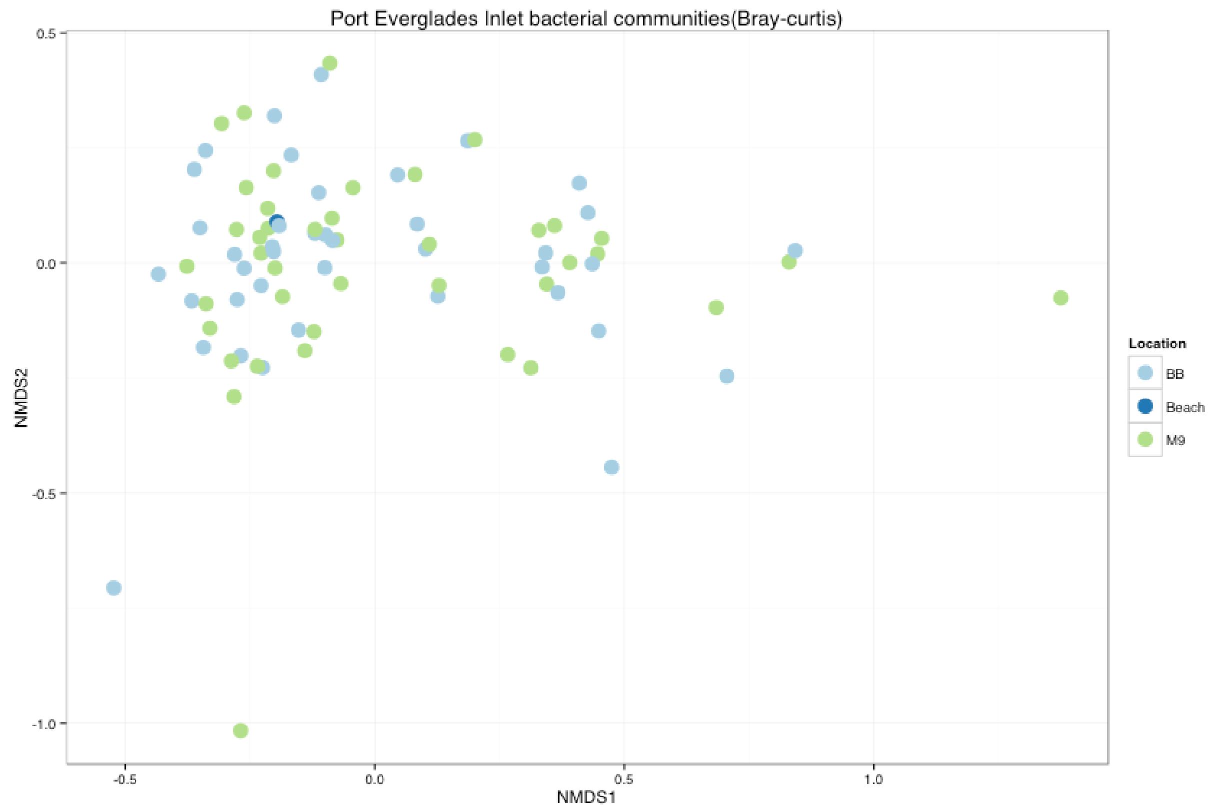


**Figure 13:** NMDS plot of Bray-Curtis dissimilarity by location

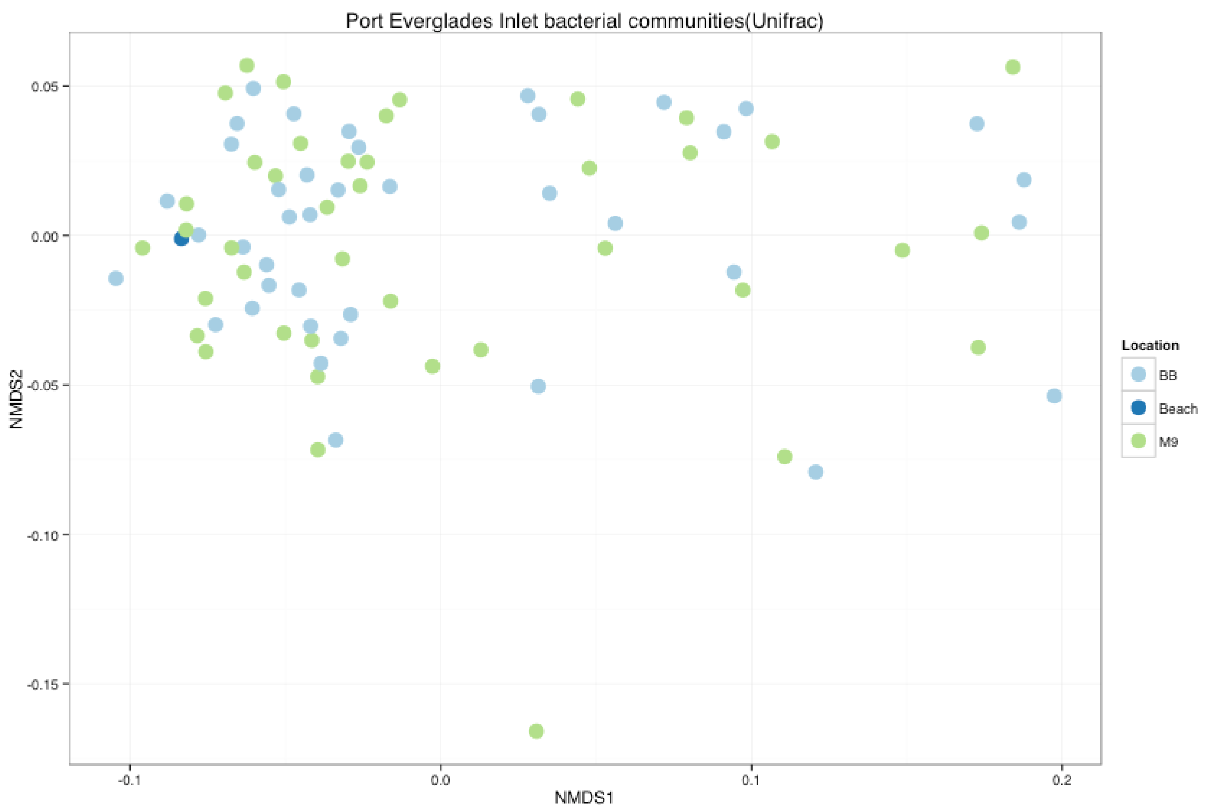


**Figure 14**: NMDS plot of weighted Unifrac distance by location

### *Season*

Southern Florida has two main seasons. The wet season, which ranges from May-September and the dry season, which ranges from October-April. Significant differences in microbial community composition were observed between seasons. There is evident clustering of samples by season in the ordination plots (Figs. 15 and 16). Adonis tests were run to assess statistical significance of changes in community composition. The results for both the Bray-Curtis dissimilarity and weighted Unifrac distance are significant ($P<0.05$, $R^2 = 0.157$ (Bray-Curtis); $P<0.05$, $R^2 = 0.203$ (Unifrac)).

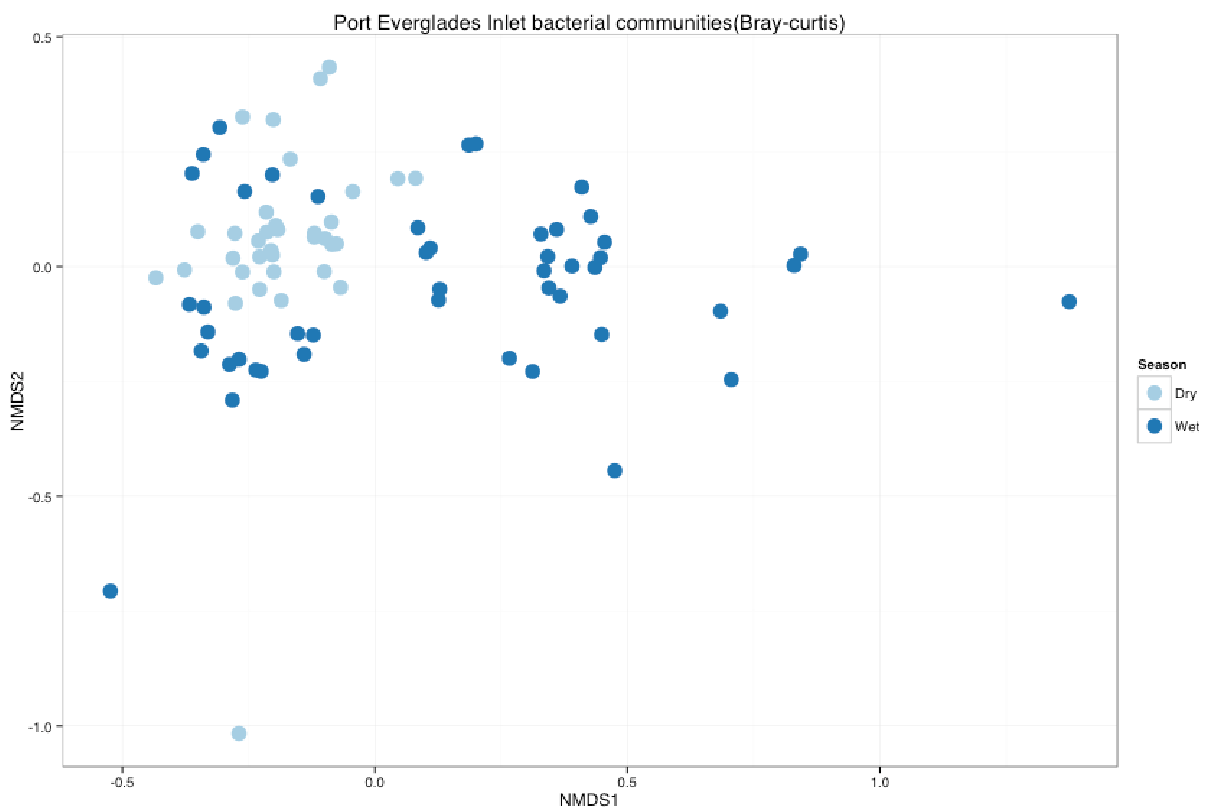


**Figure 15**: NMDS plot of Bray-Curtis dissimilarity by season

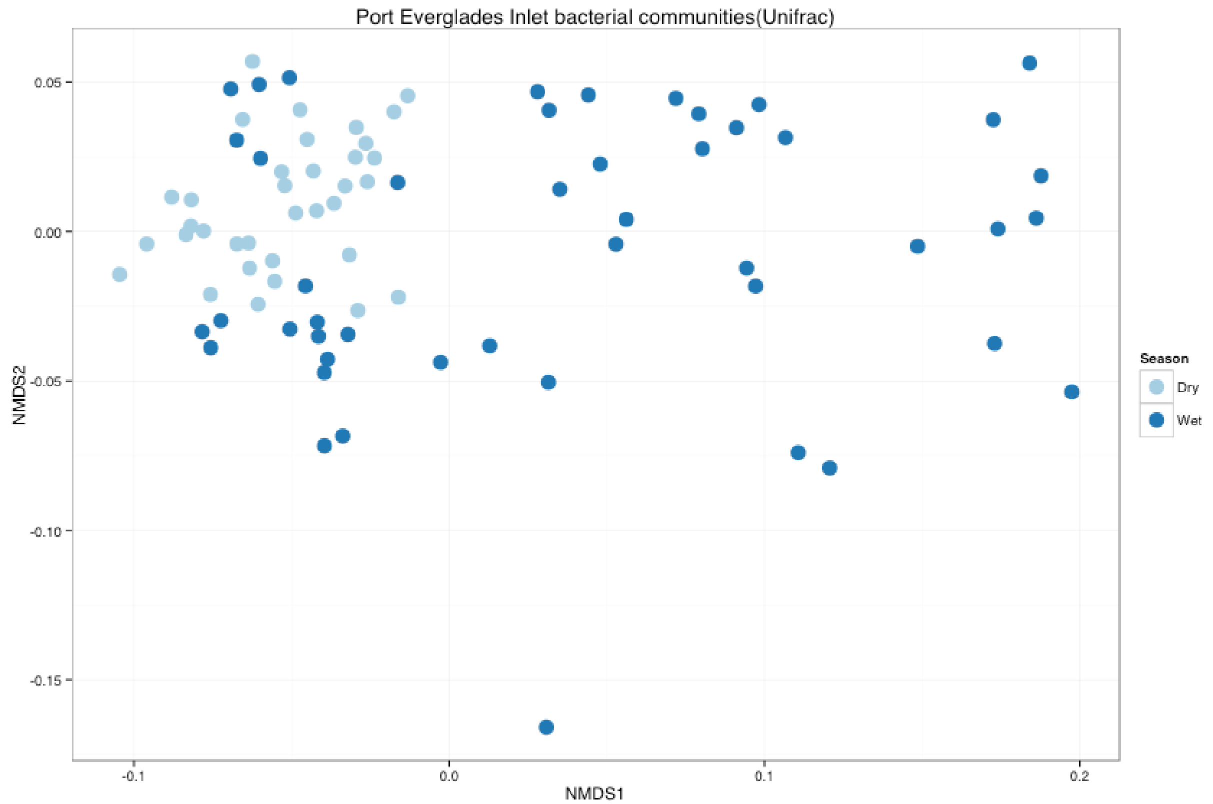


**Figure 16:** NMDS plot of weighted Unifrac distance by season

### *Month*

A major objective in this study was to analyze the Port Everglades Inlet surface water on a finer time scale then had previously been done. Water samples were taken weekly for a year to allow for samples to be analyzed by month. To determine if there were differences in the microbial community composition by month an Adonis test was run using Bray-Curtis dissimilarity values and weighted Unifrac values. The results for these tests came back significant ($P<0.05$, $R^2 = 0.605$ (Bray-Curtis); $P<0.05$, $R^2 = 0.706$)). After determining that community composition differed significantly by month, multiple pairwise comparisons between all months were done using Bray-Curtis dissimilarity values. An Adonis test was used to evaluate significant differences between pairs of months.

| Months | P-Value | $R^2$ |
|---|---|---|
| **Jan_Feb** | 0.001 | 0.231 |
| **Jan_Mar** | 0.001 | 0.341 |
| **Jan_Apr** | 0.001 | 0.358 |
| **Jan_May** | 0.001 | 0.443 |
| **Jan_Jun** | 0.001 | 0.393 |
| **Jan_Jul** | 0.002 | 0.580 |
| **Jan_Aug** | 0.001 | 0.654 |
| **Jan_Sept** | 0.001 | 0.599 |
| **Jan_Oct** | 0.001 | 0.597 |
| **Jan_Nov** | 0.002 | 0.306 |
| **Jan_Dec** | 0.001 | 0.320 |
| **Feb_Mar** | 0.001 | 0.237 |
| **Feb_Apr** | 0.002 | 0.305 |
| **Feb_May** | 0.001 | 0.365 |
| **Feb_Jun** | 0.001 | 0.238 |
| **Feb_Jul** | 0.004 | 0.476 |
| **Feb_Aug** | 0.001 | 0.568 |
| **Feb_Sept** | 0.001 | 0.475 |
| **Feb_Oct** | 0.001 | 0.453 |
| **Feb_Nov** | 0.001 | 0.281 |
| **Feb_Dec** | 0.002 | 0.292 |
| **Mar_Apr** | 0.001 | 0.350 |
| **Mar_May** | 0.001 | 0.430 |
| **Mar_Jun** | 0.001 | 0.410 |

| **Mar_Jul** | 0.001 | 0.584 |
|---|---|---|
| **Mar_Aug** | 0.003 | 0.654 |
| **Mar_Sept** | 0.001 | 0.578 |
| **Mar_Oct** | 0.001 | 0.565 |
| **Mar_Nov** | 0.001 | 0.301 |
| **Mar_Dec** | 0.002 | 0.342 |
| **Apr_May** | 0.001 | 0.195 |
| **Apr_Jun** | 0.003 | 0.308 |
| **Apr_Jul** | 0.004 | 0.552 |
| **Apr_Aug** | 0.002 | 0.612 |
| **Apr_Sept** | 0.001 | 0.556 |
| **Apr_Oct** | 0.001 | 0.553 |
| **Apr_Nov** | 0.001 | 0.344 |
| **Apr_Dec** | 0.003 | 0.421 |
| **May_Jun** | 0.001 | 0.211 |
| **May_Jul** | 0.001 | 0.546 |
| **May_Aug** | 0.001 | 0.597 |
| **May_Sept** | 0.001 | 0.567 |
| **May_Oct** | 0.001 | 0.550 |
| **May_Nov** | 0.001 | 0.380 |
| **May_Dec** | 0.001 | 0.478 |
| **Jun_Jul** | 0.015 | 0.375 |
| **Jun_Aug** | 0.001 | 0.450 |
| **Jun_Sept** | 0.001 | 0.400 |
| **Jun_Oct** | 0.001 | 0.390 |

| | | |
|---|---|---|
| **Jun_Nov** | 0.001 | 0.351 |
| **Jun_Dec** | 0.001 | 0.377 |
| **Jul_Aug** | 0.018 | 0.312 |
| **Jul_Sept** | 0.012 | 0.262 |
| **Jul_Oct** | 0.007 | 0.348 |
| **Jul_Nov** | 0.004 | 0.535 |
| **Jul-Dec** | 0.001 | 0.488 |
| **Aug_Sept** | 0.001 | 0.270 |
| **Aug_Oct** | 0.001 | 0.375 |
| **Aug_Nov** | 0.001 | 0.600 |
| **Aug_Dec** | 0.001 | 0.579 |
| **Sept_Oct** | 0.047 | 0.133 |
| **Sept_Nov** | 0.001 | 0.497 |
| **Sept_Dec** | 0.001 | 0.462 |
| **Oct_Nov** | 0.001 | 0.485 |
| **Oct_Dec** | 0.004 | 0.455 |
| **Nov_Dec** | 0.026 | 0.197 |

**Table 2**: Results of multiple comparison Adonis tests by month. Alpha= 0.01.

Results of the Adonis test reveal that all month comparisons were significant, but not at the same level. Some months show lower p-values than others, indicating these months had more differences in microbial community composition. The NMDS plots for both Bray-Curtis dissimilarity and weighted Unifrac show evident clustering of samples by month (Figs. 17 and 18). The weighted Unifrac plot shows more distinct clustering then

the Bray-Curtis dissimilarity plot. The interaction between month and location was also tested but results were insignificant (P>0.05).

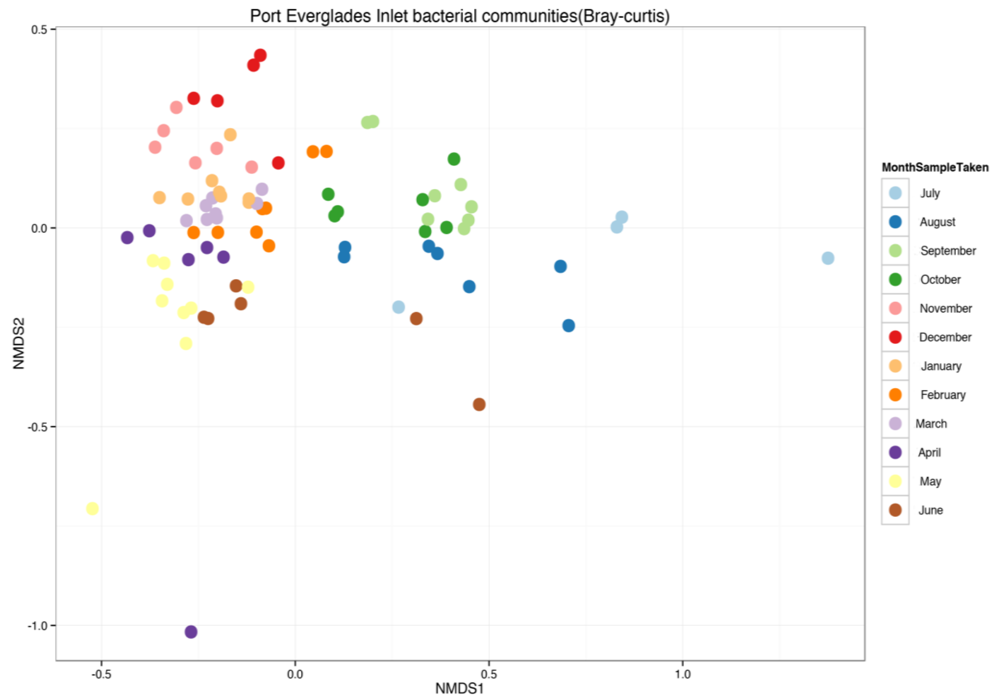


**Figure 17**: NMDS plot of Bray-Curtis dissimilarity by month

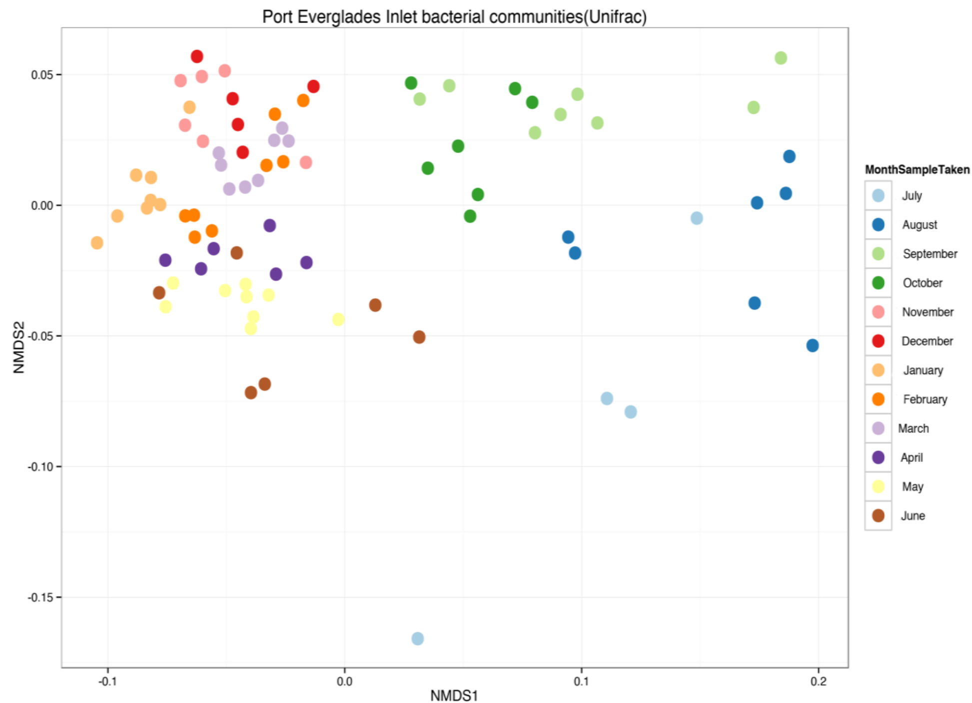


**Figure 18**: NMDS plot of weighted Unifrac distance by

**Pathogens**

Orders known to contain pathogenic organisms were partitioned out from the overall dataset and filtered to the top 100 OTUs in that class to determine how much of the community composition was made up of pathogens in water samples from Port Everglades Inlet. The major orders partitioned out from the *Firmicutes* phylum are *Bacillales, Clostridiales,* and *Lactobacillales.* The orders partitioned out from the *Proteobacteria* phylum are *Enterobacteriales, Campylobacterales,* and *Vibrionales.*

***Bacillales, Clostridiales,* and *Lactobacillales***

The top 50 OTUs in each order were plotted using a stacked bar chart of abundance level vs. month the samples were taken. The fill on the chart is genus level classification. The most common taxa in the Bacillales order contain three pathogenic

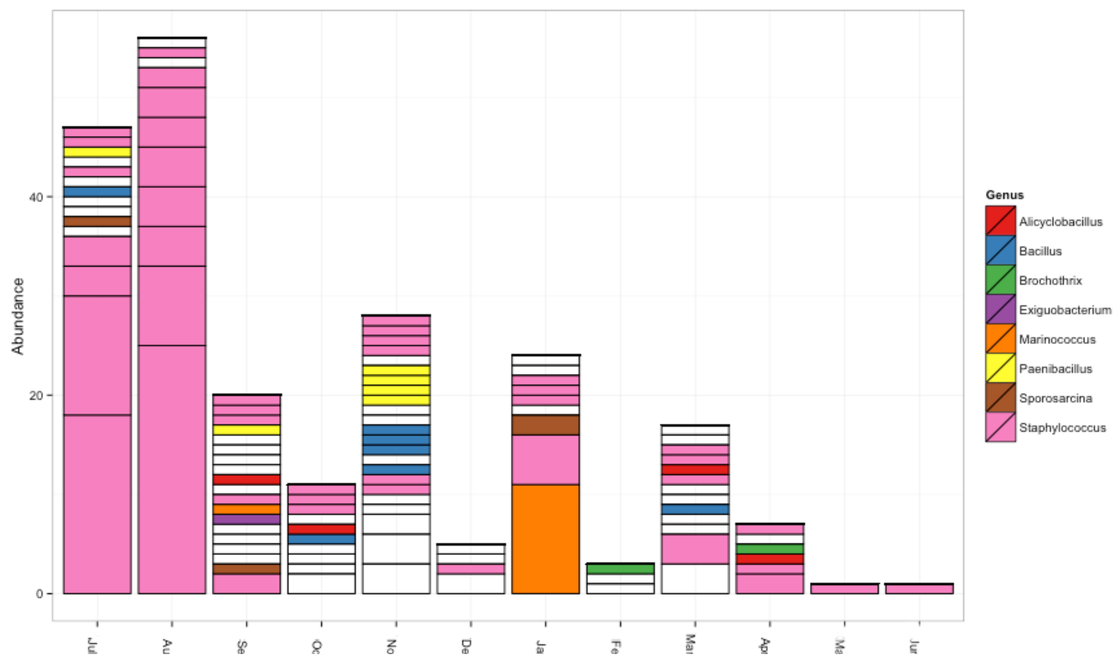


**Figure 19:** Taxa summary and abundance levels of top 50 OTUs in Bacillales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

genera, *Staphylococcus, Bacillus,* and *Paenibacillus*. Of these three genera only *Staphylococcus* and *Bacillus* contain strains which are pathogenic to humans. The presence of *Staphylococcus* is in higher abundance than any other genus in this order.

*Paenibacillus* is only present in samples taken in July, September, and November of 2013 (Fig 19). *Bacillus* is seen in July, October, November, and March. While these organisms are present in the samples, the only genus with relatively high abundance is *Staphylococcus* (Fig. 19).

The order *Clostridiales* has three classified genera in the top 50 OTUs. The only genus in this order that contains pathogenic strains is *Clostridium*. This is also the genus that is seen in the highest abundance in this order (Fig. 20).

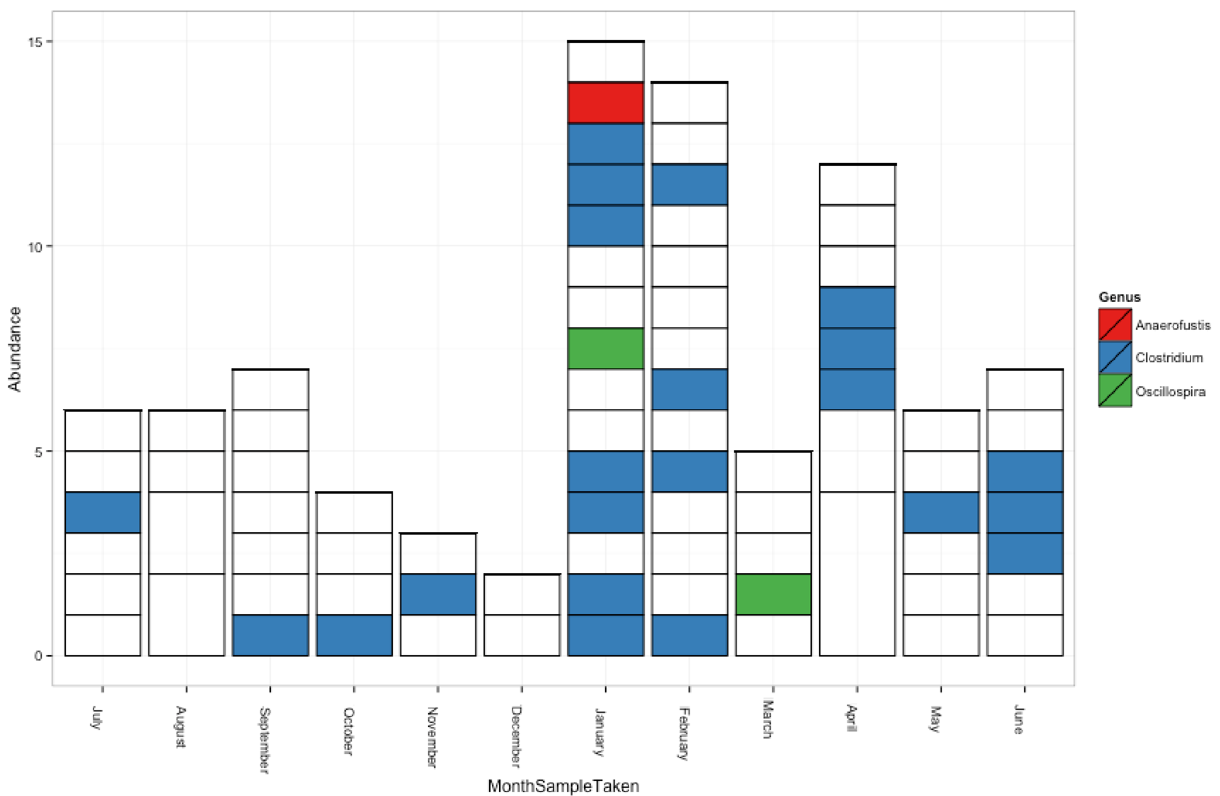


**Figure 20**: Taxa summary and abundance levels of top 50 OTUs in Clostridiales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

Of these seven genera, four are known to have pathogenic strains, *Alloiococcus, Enterococcus, Leuconostoc,* and *Streptococcus*. The most abundant genera is *Streptococcus,* it is present in all months except for November and May (Fig. 21).

*Alloiococcus* and *Enterococcus* are only present in samples from the month of September (Fig 21).

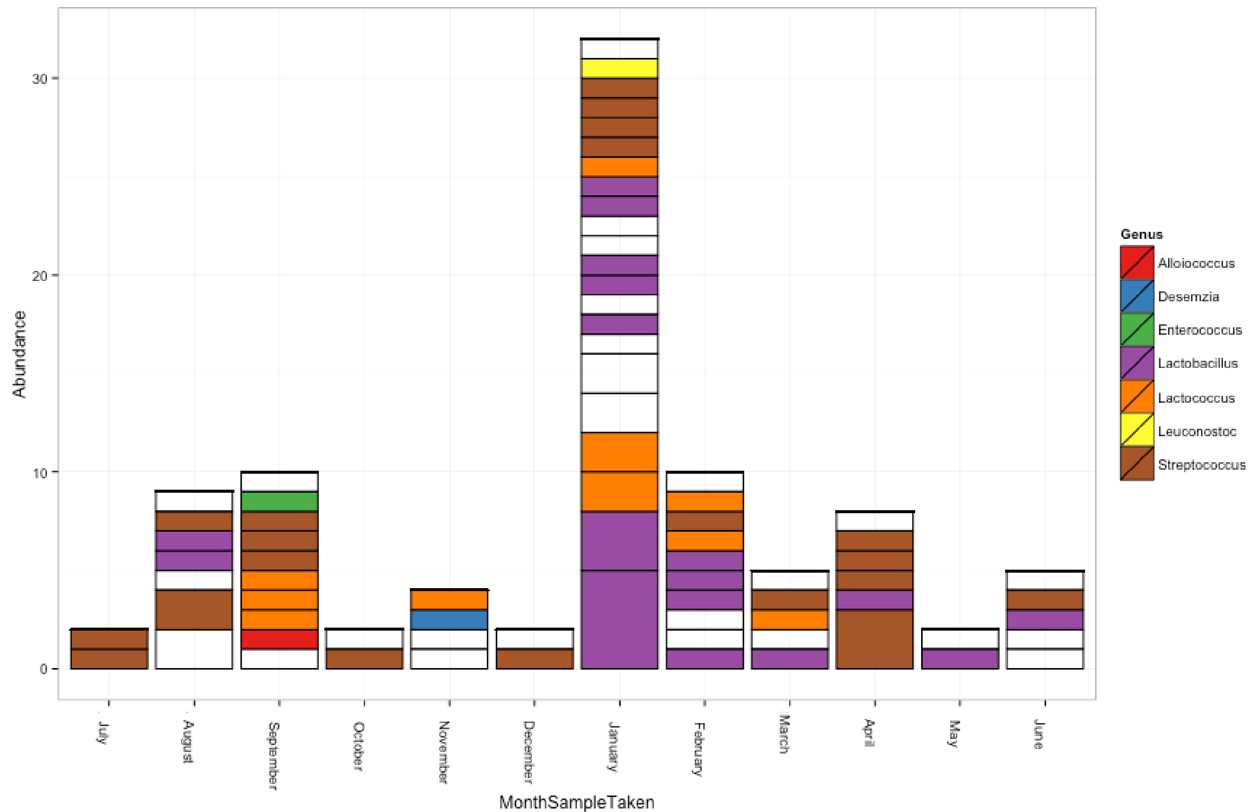


**Figure 21**: Taxa summary and abundance levels of top 50 OTUs in Lactobacillales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

### *Enterobacteriales, Campylobacterales,* and *Vibrionales*

The results of the presence of pathogenic organisms in the *Enterobacteriales* order show that of the top seven most abundant genera, five of them contain pathogenic strains, *Candidatus Phlomobacter, Citrobacter, Dickeya, Morganella,* and *Providencia*. The pathogens in these genera are mostly opportunistic and don't commonly cause

human infections. Of the five genera containing pathogenic strains, *Candidatus Phlomobacter* and *Dickeya* are plant pathogens. The most abundant genus is *Citrobacter* and appears in relatively high abundance (in comparison to other genera in this order) in all months (Fig. 22). This genus can be pathogenic in humans, but is most often not unless the

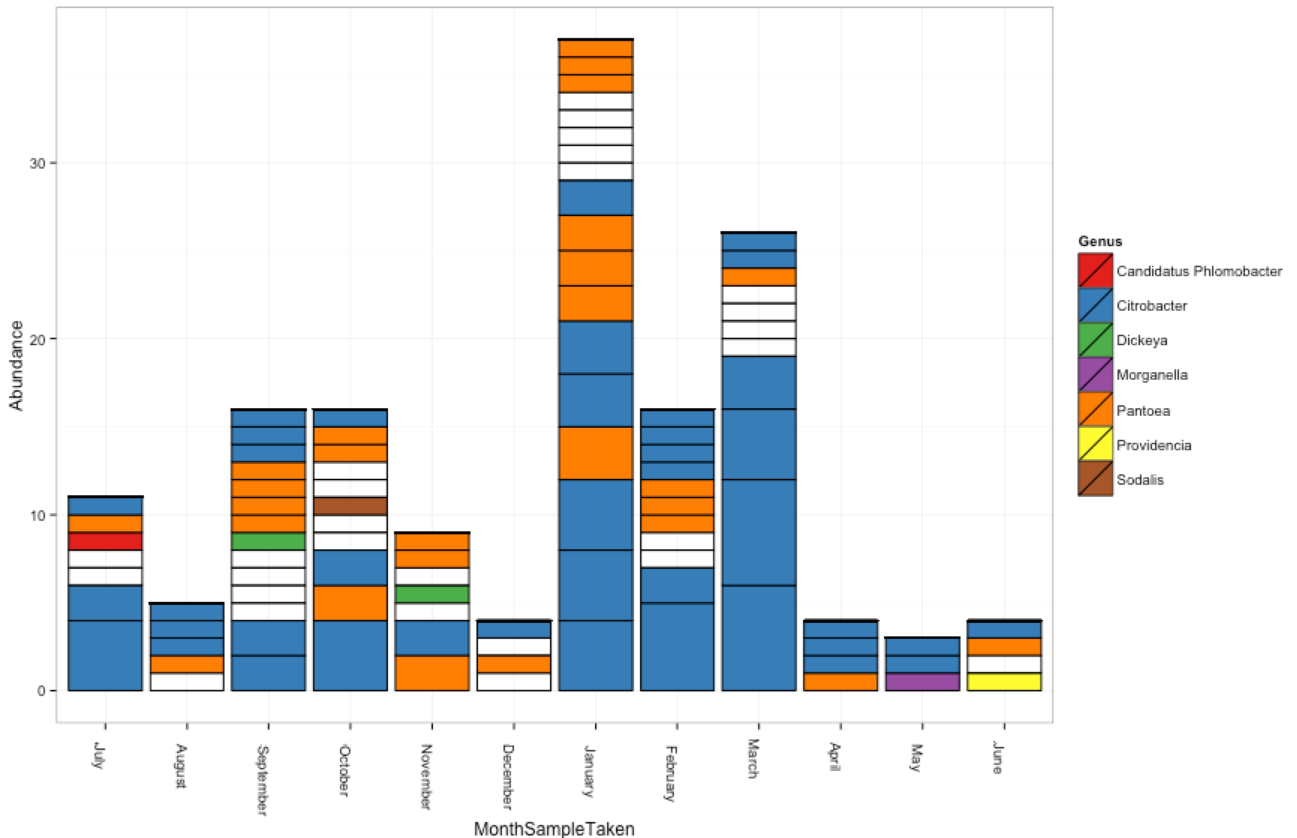


**Figure 22**: Taxa summary and abundance levels of top 50 OTUs in Enterobacteriales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

person is immunocompromised. The genus *Providencia* is an opportunistic pathogen and is only seen in the month of June (Fig. 22).

The order *Campylobacterales* only contained two classified genera in quantifiable abundance. Of these two genera only *Helicobacter* is a known human pathogen. *Helicobacter* is not present in high abundance in the Port Everglades water samples and is only seen in the month of October (Fig. 23). Overall, this order did not contain many classified organisms down to the genus level.

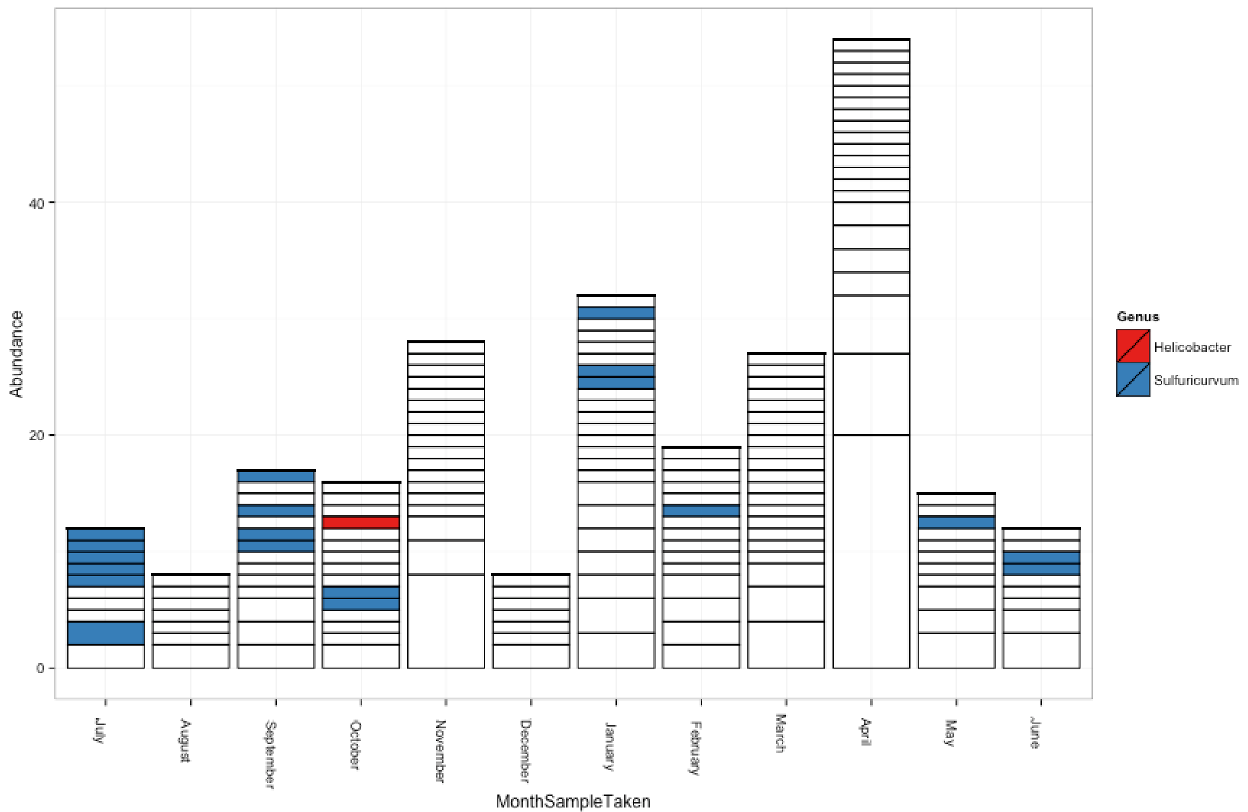


**Figure 23**: Taxa summary and abundance levels of top 50 OTUs in Campylobacteriales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

The *Vibrionales* order did not contain any known classified pathogens in the top 50 OTUs. This order was composed of the *Pseudoalteromonas* and *Photobacterium* genera. The *Photobacterium* genus was only seen in month of March (Fig. 24)

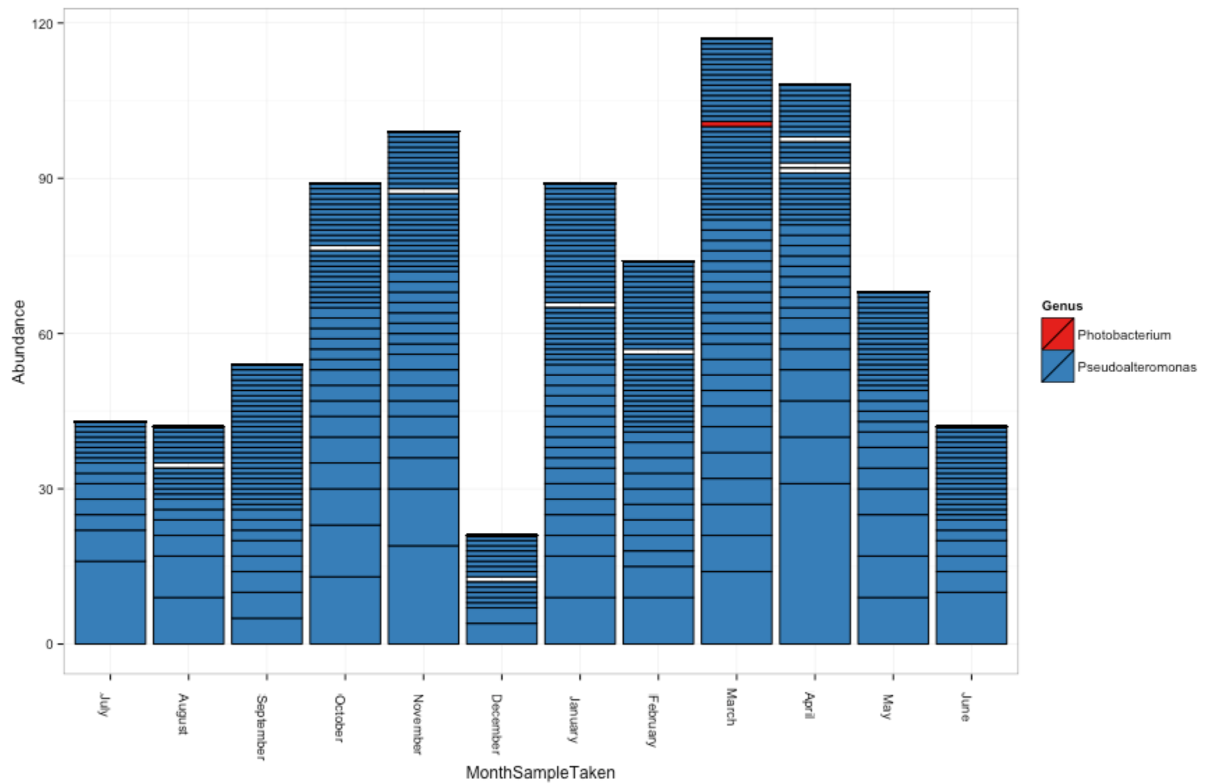


**Figure 24**: Taxa summary and abundance levels of top 50 OTUs in Vibrionales order. Used to determine the presence of pathogens. Color designations are shown on the right, while white indicates unclassified.

**Ion Concentration, Environmental Metadata Correlations, and Linear Regression Analysis:**

Environmental metadata such as salinity, rainfall, and water temperature were examined for trends against the top 100 OTUs at the phyla level. An interesting trend seen in the data is that *Proteobacteria* and *Cyanobacteria* show an inverse trend in abundances, which correlates with temperature and rainfall data to some degrees (Fig. 25).

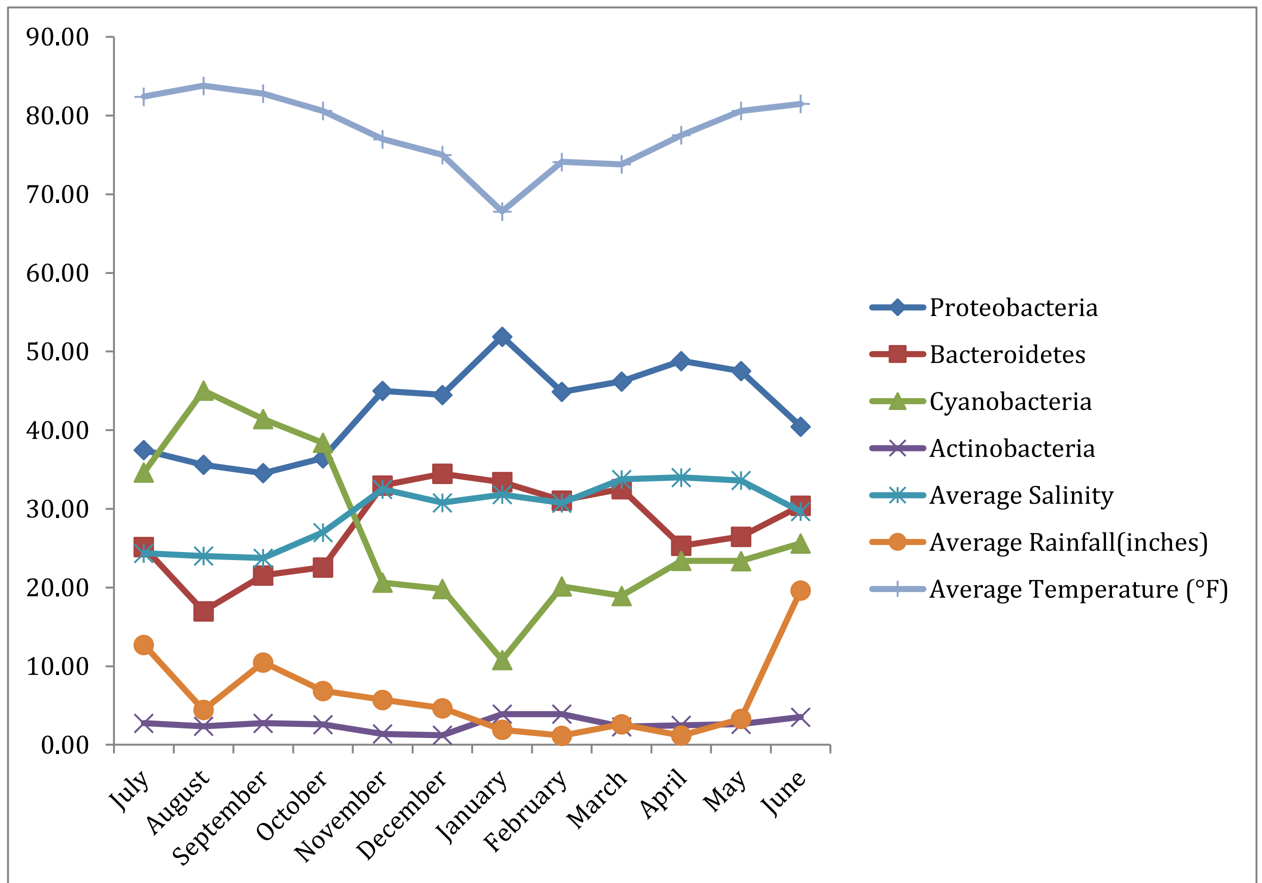


**Figure 25:** Correlation between top phyla abundance and environmental metadata: salinity, temperature, and precipitation

Multiple least Squares regression analysis was completed using the number of reads for the top nine most abundant bacterial classes and environmental metadata: chloride ion, sulfate ion, rainfall, water temperature and salinity (Table 3). The level of alpha was set to 0.10 for statistical significance. The $R^2$ value is the measure used to

determine how well the data fits the regression line. The higher the $R^2$ values the better the data fits the model.

| | Gamma | Flavobacteriia | Acidomicrobiia | Alpha | Chloroplast | Beta | Synechcoccophycideae | Actinobacteria | Unclassified |
|---|---|---|---|---|---|---|---|---|---|
| **R- Squared** | 0.21 | 0.21 | 0.21 | 0.13 | 0.13 | 0.14 | 0.20 | 0.21 | 0.07 |
| **Salinity** | 0.0173 | 0.101 | 0.0013 | 0.0008 | NA | 0.0005 | NA | 0.0514 | NA |
| **Water Temperature** | 0.0601 | 0.0204 | NA | NA | NA | NA | 0.001 | 0.0001 | NA |
| **Chloride Ion** | NA | NA | NA | NA | 0.0016 | NA | 0.0155 | NA | NA |
| **Sulfate Ion** | NA | NA | NA | NA | 0.0016 | NA | NA | NA | NA |
| **Rainfall** | NA | NA | NA | NA | NA | NA | NA | NA | 0.0184 |

**Table 3**: Results of multiple least squares linear regression analysis using SAS. Alpha =0.10. All values greater than alpha were not included. Values less than 0.10 were considered to be statistically significant. Gamma, Alpha, and Beta table headings refer to the classes Gammaproteobacteria, Alphaproteobacter, and Betaproteobacteria and were shortened for spatial reasons.

Results for the class Gammaproteobacteria ($R^2$=0.21) show a significant relationship with salinity ($p$=0.0173) and water temperature ($p$=0.0601). The class Flavobacteriia($R^2$=0.21) show a significant relationship with salinity ($p$=0.101) and water temperature ($p$=0.0204). The class Acidomicrobiia ($R^2$=0.21) show a significant relationship with salinity ($p$=0.0013). The class Alphaproteobacteria ($R^2$=0.13) show a recognizable relationship with salininty ($p$=0.0008). The class Chloroplast ($R^2$=0.13) show a recognizable relationship with chloride ion concentration ($p$=0.0016) and sulfate ion concentration ($p$=0.0016). The class Betaproteobacteria ($R^2$=0.14) show a recognizable relationship with salinity ($p$=0.0005). The class Synechococcophycideae ($R^2$=0.20) show a significant relationship with water temperature ($p$=0.001) and chloride ion concentration ($p$=0.0155). The class Actinobacteria ($R^2$=0.21) show a significant relationship with salinity ($p$=0.0514) and water temperature ($p$=0.0001). The unclassified bacterial groups ($R^2$=0.07) had a somewhat weak relationship with rainfall ($p$=0.0184).

The R-squared values are lower than expected, but at least some of the variations can be explained by the model to some extent.

## Section V: Discussion

The objective of this study was to characterize the surface water bacterioplankton community in Port Everglades Inlet on a finer sampling scale than has been previously done (Campbell et al, 2015). This is the first study, to my knowledge, that has been completed solely on Port Everglades Inlet utilizing Illumina MiSeq DNA sequencing technology. Previous to this, majority of the sampling efforts in the inlet were for human enteric pathogens, fecal indicator bacteria, and nutrient monitoring (Futch *et al.,* 2011; Collier *et al.,* 2008; Craig unpublished, 2012). Data from this study can be used in conjunction with previous and future monitoring efforts in the port to provide a comprehensive overview of Port Everglades Inlet surface water microbiome and provide a baseline for comparison in future years.

### *Port Everglades Inlet and Sample Collection*

Port Everglades Inlet was chosen as an interesting study site for this project because it is one of the busiest cruise ship and cargo ports in the country and is constantly under human influence via cruise, cargo, and recreational ship traffic. The inlet has constant boat traffic with the total number of ships called into the port in 2013 exceeding 3,850 and increasing to over 3,970 in 2014. These numbers do not include any personal or recreational boats entering the inlet that do not dock there. The total number of passengers that entered the port via cruise ships was 3,600,636 in 2013, rising to 4,001,354 in 2014. The total number of barrels of petroleum brought into the port in 2013 was 109,080,601 increasing to 112,370,083 in 2014. The ports total revenue for both 2013 and 2014 were record highs generating over $142.9 million dollars in 2013 and over $146.8 million dollars in 2014 (Port Everglades Commerce Report, 2013; Port Everglades Commerce Report, 2014). Port Everglades is a major source of income for Broward County, but is also a major source of pollution into the surrounding marine and coastal environments (Banks *et al.,* 2008; Carrie Futch *et al.*, 2011; Walker *et al.*, 2012). Therefore, examining fluctuations of the inlet's microbiome throughout the year provides

valuable baseline data to help manage and maintain the health of the inlet ecosystem as well as the coastal areas and marine environments adjacent to the inlet.

Surface water samples were collected in 1 L glass Pyrex bottles in Port Everglades Inlet. Water samples were filtered within 30 minutes of collection using a vacuum pump and 0.45µm pore size filter. The filter size of 0.45 µm is not ideal as the size range for most bacteria is from 0.2-0.5 µm, therefore, some organisms may have passed through the filter (Water Quality Association, 2015). For future studies a 0.1-0.2 µm filter is recommended to ensure that no microbes pass through the filter. A recent study completed by Padilla and colleagues (2015) concluded that both pore size, and amount of water filtered can drastically bias microbial community diversity in 16s amplicon studies. They suggested that filter fractionation of samples should only be used for quantitative comparisons if variations in sample volume are shown to be negligible (Padilla *et al.,* 2015). In this study the same sample volumes and pore sizes were used for all samples, and results are therefore considered accurate.

### *Alpha Rarefaction Curves as a Measure of Sampling and Sequencing Efforts*

Rarefaction curves are generated to determine if sampling depth was sufficient to capture full species diversity in the samples. In this study rarefaction curves were generated in both QIIME and phyloseq. The results for both curves were similar. The rarefaction curves for Shannon Diversity, which accounts for community richness and evenness, completely plateau meaning that it is likely that full species diversity was captured for all samples at the current sampling effort. In contrast to this, the rarefaction curves for Observed species and Chao1, which are species richness estimators and account for the presence of singletons, do not reach a complete plateau, even at the highest number of reads (80,122). This indicates that sampling efforts were not saturated. The reason that these two indices show contrasting results is based on how the indices are calculated. When calculating the Shannon Index the equation uses the proportions organisms in the sample. If there are a lot of counts for one organism it is likely that this organism will be randomly sampled more times than something without as many counts,

no matter what the sequencing depth is set too. If sequencing depth is increased for the rarefaction, the overall proportions of the organisms will remain relatively the same making the Shannon Index a relatively stable measure, generating curves that reach a plataue. The Chao1 and Observed Species Indices differ from the Shannon Index because the curves for these indices increase as rarefaction depth increases. This occurs because these measures account for singletons, or unique taxa, present in the samples. Consequently, if only one count of a specific organism is obtained from sequencing, if the sequencing depth is set deep enough, the rarefaction will find it increasing the number of observed species, and accordingly increasing the slope of the curve.

In this study, when including all samples, full sequence diversity was not captured because equal sequencing depth was not reached in all samples. While the proportions of different taxa in the samples remain relatively stable, samples with higher number of reads had more unique taxa, resulting in increased species richness. In future studies, resequencing of samples with a low number of reads in comparison to other samples can help to reach equal sequencing depth and capture full sequence diversity.

***Rank Abundance Curve***

The rank abundance curve at 97% sequence similarity shows that the Port Everglades Inlet surface water samples are relatively high diversity samples. The slope of the curve is gradual with a longer tail indicating the presence of singletons, increasing diversity of the samples. To ensure that these singletons are not sequencing artifacts chimera checking was completed on all sequences. The front-end of the curve indicates that a few OTUs are dominant in the samples. The two OTUs with the highest rank abundance belong to the family's Rhodobacteraceae (2717) and Cryomorphaceae (1389). Both of these organisms have been found in high abundance in previous coastal water and inlet studies (Campbell *et al.,* 2015; Baltar *et al.,* 2015).

***Comparison of Alpha Diversity Indices***

Alpha diversity indices were calculated to determine community diversity and community richness for Port Everglades Inlet water samples. The results for Observed Species, Chao1, Shannon, and Simpson diversity indices showed relatively similar levels

of diversity between all months, with no significant differences. The Shannon and Simpson diversity indices only represent species richness, and do not account for abundance of organisms. The Chao1 and Observed Species indices account for both species richness and abundance measures. The months of November, January, and April showed high abundance and high diversity levels, with November and January also being very high in species richness. The month of July shows very high species richness but low species abundance. The month of August was the lowest in species richness, abundance, and diversity levels. A previous study completed on Port Everglades Inlet showed similar diversity estimation values except for in the month of April, where low species richness was observed (Campbell *et al.,*2015).

***Microbial Community Composition Taxa Fluctuations*: *Location, Month, and Season***

The most abundant phyla in all samples (>1%) were Proteobacteria, Cyanobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia, and Euryarchaeota. These organisms are consistent with previous studies completed on marine coastal waters (Campbell *et al.,* 2015; Gifford *et al.,* 2014; Elifantz *et al.*, 2013; Rappe *et al.,* 2000). Significant differences in community composition were seen in alpha diversity when comparing the month of August with the months of December, October, and November. This is most likely due to the low species richness, diversity, and abundance seen in the samples for the month of August. The low species richness, diversity, and abundance values in August, could be due to the abnormally low rainfall for this month. Beta diversity compares differences among groups. In this study the groups examined were location, month, and season. Significant differences were observed only when comparing the groups for month and season. Location was excluded as a variable from any further analysis. At the phylum taxonomic level only slight fluctuations in microbial community composition can be seen throughout the year. The most drastic shifts occur with Cyanobacteria, which decrease in relative abundance during the winter months, or dry season, and increase in abundance during the summer and early fall, or wet season months. My data correlates with previous observations of increased cyanobacterial blooms in Florida's coastal and freshwater ecosystems in the late summer and early fall months. The blooms are caused by warm water conditions paired with increased sunlight

levels, and nutrient loading from urban runoff (Flombaum *et al.,* 2013; Florida Department of Environmental Protection, 2011). . At the class taxonomic level the most common taxa were Alphaproteobacteria, Flavobacteriia, Synechoccophycideae, Chloroplast, Gammaproteobacteria, Actinobacteria, and Acidimicrobiia. These results are similar to taxonomic composition seen in other coastal microbial community studies (Campbell *et al.,* 2015; Gifford *et al.,* 2014; Elifantz *et al.*, 2013; Rappe *et al.,* 2000). More pronounced fluctuations in microbial community composition are observed at the class taxonomic level, specifically with Synechoccophycideae, in which relative abundance doubles during the wet season months, while Flavobacteriia decreases during that same time. Synechoccophycideae is the most abundant organism in the Cyanobacteria phylum, and therefore would be expected to have a similar shift in abundance to what was seen at the phylum level. The class Chloroplast, which is also in the Cyanobacteria phylum, was abundant in the samples. Interestingly, this class shows no noticeable monthly or seasonal shifts, despite the overall fluctuations seen in the Cyanobacteria phylum. The most pronounced shifts in microbial composition can be seen at genus level classification. The most abundant organisms at the genus level were *Candidatus portiera* and *Synechococcus*. These two organisms were also seen in the highest abundance in a previous study completed on Southeast Florida's inlets, outfalls, and reef environments (Campbell *et al.,* 2015).

### ***Significance of Abundant Taxa and Correlation with Environmental Metadata***

The most abundant bacterial phyla seen in this study were Proteobacteria, Bacteroidetes, Cyanobacteria, and Actinobacteria. The most abundant classes were Alphaproteobacteria, Flavobacteriia, Synechoccophycideae, Chloroplast, Gammaproteobacteria, Actinobacteria, and Acidimicrobiia. Bacterioplankton in these classes have been shown to dominate many coastal and oceanic waters (Campbell *et al.,* 2015; Gifford *et al.,* 2014; Williams *et al.,* 2013; Elifantz *et al.*, 2013; Rappe *et al.,* 2000). Within these taxonomic ranks, there are many microbial organisms that play important roles in biogeochemical, energy, and nutrient cycling in the oceans (Georges *et al.,* 2014; Gonzalez and Moran, 1997). Therefore it is important to examine the presence

of these organisms in an ecosystem known to be a point source of pollution to surrounding marine environments.

*Alphaproteobacteria*

Alphaproteobacteria have been shown to dominate the world's oceans surface water microbial habitats (Biers *et al.,* 2009; Giovannoni and Stingle, 2005). Abundant microbial members observed from this class in this study were Rhodobacterales, Pelagibacteraceae, Pseudonana, and Rhodospirillales. The most dominant organism being members of the Rhodobacterales order. Members of this order have been documented to be rapid primary surface colonizers in Atlantic Ocean coastal waters and have also been observed as primary biofilm colonizers in subtropical and tropical coastal waters (Elifantz *et al.* 2013; Dang *et al.,* 2008). Previous studies investigating the responses of bacterioplankton to additions of anthropogenic stressors (nutrient enrichment and acidification) showed increased abundance of *Rhodobacteraceae* in response to increased nutrient availability. This study was completed by season, showing an increased abundance of *Rhodobacteraceae* from early spring into summer months. While data in this study reflects an increased abundance of Rhodobacterales in the early spring months, this group is also in high abundance during some of the winter months. The increased abundance of this group may be caused by peak shipping season, which runs from November to April, and may increase the amount of anthropogenic influences introduced into the inlet waters. Another possible reason in which this increase in abundance may be seen in these months could be due to the absence of Rhodospirillales. In a paper by Ahrendt and colleagues(2014) it was observed that a sharp drop in the abundance of Rhodobacterales was seen in response to an increased abundance of Rhodospirillales. These bacteria are purple-sulfur phototrophs and are often enriched in hypersaline habitats (Ahrendt *et al.,* 2014).

The heterotrophic microbes of the family Pelagibacteraceae are also present in high abundance in Port Everglades Inlet samples. This group is also known as the

bacterial clade SAR11 (Kutschera and Schauer 2012; Brown *et al.,* 2012). The SAR11 clade is considered one of the most successful microbial assemblages on the planet. This group of microbes is found universally in every ocean, in both coastal and open ocean environments, and comprises about 35% of ocean surface water microbial communities (Morris *et al,* 2002; Brown *et al,* 2012). Studies examining the SAR11 clade have found that different phylotypes inhabit different geographical locales and this is driven by physical parameters such as temperature, latitude, and salinity. In coastal environments this group has been observed to increase in abundance during the warmer summer months when tropical waters are highly eutrophic and there is high UV radiation (Brown *et al,* 2012). This observation is supported by my data. Increases in abundance of Pelagibacteraceae are observed in the spring and summer months, when water temperatures began to rise and large UV radiation is at its highest.

*Gammaproteobacteria*

The most abundant organism found in the Gammaproteobacteria class was *Candidatus portiera.* This organism is an obligate primary endosymbiont whose host is whiteflies (Jiang *et al,* 2012). While it is possible that this bacterium may have been introduced into coastal waters from insects, as *Candidatus portiera* has been seen in other oceanic studies, it is most likely not a core organism in the inlet's surface water microbiome. A possible explanation is that the Greengenes database is out of date, and this organism had at least 97% sequence similarity and was therefore classified as the whitefly endosymbiont. *Candidatus portiera* is in the order Oceanospirillales. This order is mostly comprised of marine bacteria which degrade complex organic compounds, using ammonium, alkanes, and polycyclic aromatic compounds as their energy source (Jenson *et al,* 2010). In the order Oceanospirillales there is a clade of bacterium that are strictly marine invertebrate-associated microbes which are also in the *Candidatus* genus. This bacteria is known to inhabit the Atlantic Ocean and is also associated with coral reefs (Jenson *et al¸* 2010). It is possible that the *Candidatus portiera* hits were a part of the marine invertebrate-associated clade of organisms in the Oceanospirillales order. Greengenes is only able to classify organisms based on what DNA sequences are currently known, and if DNA sequences for these organisms are unknown or the database

is not updated, classification error may occur (Campbell *et al,* 2015). *Candidatus portiera* increased in abundance in the late winter months and early spring and summer months and drastically declined in abundance in the later summer and fall, reasons behind this are unknown.

Members of Alteromonadales were present in relatively high abundance in the Inlet water samples. This order is comprised of marine, Gram-negative bacteria. This group is known to be common in marine surface coastal waters in both the Atlantic and Pacific Oceans. They are common colonizers of biofilms, playing important roles in their formation. They also play important roles in coastal waters by helping to colonize the surface water microbiome in these environments (Dang and Lovell, 2000; Jones *et al,* 2007; Huang *et al,* 2008; Dang *et al,* 2008; Dang *et al*, 2011). Alteromonadales increased in abundance in the fall and winter months. Although pH was not tested in this study, previous research has examined microbial community shifts in response to season and pH levels. At decreased pH levels shifts in coral-associated bacterial communities have seen increased levels of organisms in the Alteromonadales order as well as other Gammaproteobacteria (Krause *et al*, 2012; Meron *et al*, 2012). Increased abundance of Alteromonadales was seen during fall and winter months (Krause *et al,* 2012). My data supports increased abundance of Altermonadales in the fall and winter months.

Results indicate that there is a significant relationship between Gammaproteobacteria and water temperature and salinity. Previous studies have shown that there is a relationship between Gammaproteobacterial growth and temperature (Bretter *et al,* 2003 and Lefort and Gasol, 2013). Both studies showed an increase in Gammaproteobacterial growth with increased temperatures, with an optimal growth temperature ranging from 30-40ºC (Bretter *et al,* 2003). In this study, increased growth and abundance of this class was seen to range from the winter months (December-February) to the early summer months (March-June). This is interesting as water temperatures were highest in the late summer months (July and August), when abundance levels of gammaproteobacteria decreased. It is unknown why abundance levels of this class increased in the winter-early summer months. It is possible that during this time

there were increased nutrient levels and somewhat higher temperature which may have stimulated bacterial growth.

Salinity showed a significant relationship with Gammaproteobacteria abundance in Port Everglades Inlet. A previous study showed that there is a relationship between Gammaproteobacteria and salinity, although the extent of the relationship is unknown and is most likely influenced by other environmental factors such as light availability (Campbell and Kirchman, 2012).

*Betaproteobacteria*

The only Betaproteobacteria organism seen in the top 100 OTUs is from the order Methylophilales and the OM43 clade. This clade of bacteria was first described in 1997 and was found in samples from the US Atlantic Coast. Until 2012, no known microbes from this group were cultivable (Huggett *et al,* 2012). The first cultured strain in this class was HIMB624. This clade of organisms is considered methylotrophic bacteria, meaning that they can metabolize and assimilate one-carbon (C1) compounds including, methane, methanol, methyl halides, and methylated sulfur compounds. In previous studies it has been observed that methylotrophs are associated with phytoplankton blooms. In the oceanic surface waters phytoplankton growth has been associated with the production of methanol, methylamines, and methyl halides. These compounds are used as energy sources for methylotrophs (Neufeld *et al,* 2008). The order Methylophilales was relatively stable in abundance throughout the full year. Slight increases in abundance were observed in the winter months. While this is not what would have been expected if co-occurring with phytoplankton blooms (most often seen in the summer months), the rainfall and temperature values in South Florida in the winter months were warmer and wetter than previous years, with temperatures being exceptionally warmer than normal. This may have provided optimal growth conditions for many bacterial groups.

Betaproteobacteria showed a significant relationship with salinity. An increased abundance of this taxa was seen in the winter months when salinity levels were increased in comparison to other months. In a previous study by Jackson and Vallaire (2009), the responses of microbial assemblages to salinity and nutrients in wetlands were assessed.

This study determined that Betaproteobacteria increased in abundance in response to increased salinity levels, and slightly decreased in abundance in response to increased nitrogen levels. The decreased abundance of this taxa in response to increased nitrogen levels is because Betaproteobacteria are oligotrophs and can live in habitats with relatively low nutrient levels (Jackson and Vallaire, 2009 and Mitsui *et al,* 1997). While this study determined there was an increase in abundance of this taxa with salinity, the relevance of salinity to this group is unknown due their high ecological diversity (Jackson and Vallaire, 2009).

*Flavobacteria*

The most abundant group of bacteria in the Bacteroidetes phylum was from the Flavobacteriales order. Within this order, two families were present in high abundance over all months, Cryomorphaceae and Flavobacteriaceae. Abundance levels of members classified as Cryomorphaceae were highest in the month of July, but also showed high abundance in December. High abundance levels for the month of July make sense due to increased water temperatures and UV radiation available for growth. The linear regression analysis for this study showed that there were significant relationships between this taxa and water temperature. Increased abundance in the month of December could be due to the high amount of rainfall and warm water temperatures. Increased water temperatures and high rainfall can increase the amount of runoff into the inlet, leading to nutrient enrichment in the water. These conditions could lead to increased growth of bacterial groups within the inlet. In a previous study completed on Port Everglades Inlet, Cryomorphaceae was found in high abundance in inlet samples (Campbell *et al,* 2015). A study completed by Ghiglione and Murray (2012) observed increased abundance of Cyromorphaceae in the summer months in Antarctica. The increased abundance was most likely influenced by phytoplankton bloom events, and nutrient enriched waters.

Members of the Flavobacteriaceae family were present in the highest abundance in March, April, and May. This bacterial group is considered a major component in all ocean microbial communities. They are important organisms in the microbial loop, breaking down large organic molecules such as chitin and proteins (Tully *et al,* 2014).

They are also associated with areas of high primary productivity and can break down algal polymers (Gomez-Pereira *et al,* 2010). In this family, is the genus *Psychroserpens*. This genus has been identified in reef communities in Thailand and has been associated with amoebic gill disease in fishes. In previous studies this genus was found present year-round, with increases in abundance in the summer. In my data, the genus *Psychroserpens* was present at its highest abundance in the months of February, March, and May and was almost absent in the month of September, and in very low abundance in October and August.  The low abundance of this organism in August, September, and October could be due to the fluctuations in precipitation. The month of August had significantly low precipitation and increased salinity values. In comparison, the month of September had very high precipitation values and low salinity. These drastic fluctuations may have impacted the survivability of *Psychroserpens sp.*

*Chloroplast*

Members of the Chloroplast class, in the Cyanobacteria phylum, were seen in high abundance across all months. Within this class, organisms from the orders Stramenopiles, Cryptophyta, and Haptophyceae were present in the highest abundance. Stramenopiles were seen at their highest abundance levels in the months of August and September, while members of Cryptophyta are seen in highest abundance in December and March. Stramenopiles are single-celled bacterial grazers which are present in all oceans. These organisms are a part of the marine stramenopiles (MAST) group comprised of multiple clades (Thaler and Lovejoy, 2014). A study completed on marine stramenophiles in the North Pacific Ocean saw increased abundance during the summer months when water temperatures were warmest and precipitation levels were highest. The warm water temperatures and increased freshwater runoff into the coastal environment provided ideal conditions for bacterial growth (Lin *et al,* 2012). Data from this study follows a similar pattern, with the highest abundance levels in the South Florida's rainy, summer months. Results from the linear regression analysis also showed a significant relationship between the Chloroplast group and chloride and sulfate ion concentration. Members of the Chloroplast group have been shown to carry out complete sulfate reduction and this reaction is mostly light-dependent (Schwen *et al,* 1976).

Members of Cryptophyta are classified as eukaryotic algae. These organisms are common in oligotrophic waters and can be photosynthetic or heterotrophic (Bridoux *et al,* 2015). Due to the oligotrophic nature of Cryptophyta it is not unusual that these organisms were seen to increase in abundance in the dry season months when nutrients in the port are more limited than in the summer wet season months.

*Synechoccophycideae*

*Synechococcus* and *Prochlorococcus* were the two most abundant groups in the Synechoccophycideae class. *Synechococcus* is seen in much higher abundance than *Procholorococcus* over the entire year. *Synechococcu*s was seen in its highest abundance in the summer months or the wet season, with large drops in abundance during the winter months. The opposite was seen in *Prochlorococcus*, with its highest abundance being in the winter, or dry season months. These two genera make up the majority of picoplanktonic marine cyanobacteria and have the ability to acquire important nutrients at micromolar concentrations (Palenik *et al,* 2003). Of the two genera, *Synechococcus* is usually less abundant in oligotrophic waters, but is more common in marine waters worldwide (Palenik *et al,* 2003). *Procholorococcus* was seen in higher abundance in the winter months, which is Florida's dry season and typically has lower nutrient availability than in the wet summer months. As stated previously, the winter months for 2013-2014 were unusually warmer and wetter than previous years possibly creating optimal growth conditions for more bacterial groups. The winter months are also prime shipping season in Port Everglades Inlet. All of these factors could have contributed to increased runoff and nutrient availability in the inlet's waters leading to a nutrient-rich environment, and therefore stimulating bacterial growth in the inlet during the winter months.

*Acidomicrobiales*

The order Acidomicrobiales was the highest represented member in the Actinobacteria phylum. While these organisms were not present in as high abundance as seen in other organisms, they were present in greater than 1% in all samples. The most common organism seen in this order was the strain *OCS155*. Not much research has been completed on this strain. A study by Needham and colleagues (2013) did find that this

strain fluctuates drastically with time, decreasing and increasing dramatically in abundance levels over the course of 10 days (Needham *et al,* 2013; Campbell *et al*, 2015).

*Planctomycetia*

The class *Planctomycetia* was detected in water samples in Port Everglades Inlet, but were in relatively low abundance year-round. Within the *Planctomycetia* is the genus *Planctomycetes. Planctomycetes* are facultative aerobes and anaerobes and are major denitrifiers in the oceans. *Planctomycetes* are associated with polluted waters as they are chemoorganotrohic(Siniscalchi *et al,* 2015).In this study *Planctomycetes* were seen in increased abundance ranging from October-February. The increased abundance in this group may have been caused by prime shipping season in Port Everglades Inlet. Increased boat traffic in the port leads to more anthropogenic influences into the inlet waters. In addition to increased boat traffic, the winter months in southern Florida for 2013-2014 had relatively high rainfall levels and increased water temperatures. All of these factors may have led to increased nutrient runoff, creating a high nutrient and high temperature environment, which may have led to optimal growth conditions for bacterial organisms

*Verrucomicrobiae*

In the *Verrucomicrobia* phylum, the most abundance class was the *Verrucomicrobiae*. *Verrucomicrobiae* are a relatively new taxonomic division in the bacterial domain, only being fully described in 2006 (Schlesner *et al*). They are believed to be in close relationship to the Planctomycetes group. They are ubiquitous in both aquatic and terrestrial habitats and are believed to be associated with polluted and eutrophic areas (Freitas *et al,* 2012; Schlesner *et al,* 2006). To date, there is still little known about the biological roles that *Verrucomicrobia* play in the environment. In this study *Verrucomicrobiae* were present in low abundance year round, with no distinct increases or decreases in abundance. Previous studies completed in the Lopez lab have also shown the presence of *Verrucomicrobiae* in both water samples and sponge samples (Campbell *et al,* 2015; Mulheron MS thesis, 2015 Unpublished). Future research on this

phylum needs to be completed in order to understand the roles this bacterial taxa plays in marine environments.

The overall bacterioplankton community composition seen in this study was similar to other coastal water microbial community studies (Campbell *et al.,* 2015; Gifford *et al.,* 2014; Williams *et al.,* 2013; Elifantz *et al.*, 2013; Rappe *et al.,* 2000). This is especially true when comparing this study to a previous study completed on Port Everglades Inlet microbial community (Campbell *et al,* 2015). An interesting trend seen in my dataset shows increased abundance of many bacterial groups during the dry season months of southern Florida. This was unexpected as most bacterial communities are known to proliferate in the wet summer months when sunlight and precipitation levels are at their highest. A reason why an increase in abundance of many of these organisms was seen in the winter months could be due to the unusually warm and wet, dry season southern Florida experienced in the 2013-2014 year. This also coincided with prime shipping season for Port Everglades, which runs from November to April. Higher than normal precipitation and temperature levels could lead to increased runoff into the inlet waters, leading to eutrophication. Addition of anthropogenic impacts from the shipping industry could also contribute to eutrophication of inlet waters. All of these factors combined supply the bacterial community with prime growing conditions, and could therefore explain the increased microbial abundance seen during this time.

***Presence of Pathogens in Port Everglades Inlet Surface Waters***

Port Everglades Inlet is known to be a point source of pollution introducing harmful pollutants into the surrounding marine environments including a coral reef tract and recreational beaches (Banks *et al.,* 2008 and Stamates *et al.,* 2013). Due to the influence the inlet waters have on the surrounding marine environments it is important to examine the presence of pathogenic organisms in the inlet waters. The presence of pathogens in inlet water samples were determined by partitioning out the top 100 OTUs of bacterial orders known to contain human and animal pathogens relevant in the surrounding environments.

*Bacillales*

The order Bacillales contained three known pathogenic genera *Staphylococcus, Bacillus,* and *Paenibacillus*. The only one of these genera known to cause ocean-related illness is *Staphylococcus spp. Staphylococcus* was present in almost all months and had the highest abundance levels of all three genera. *Staphylococcus* abundance levels were highest in the months of July and August. *Staphylococcus* is a genus of gram-positive bacteria commonly found on the nails, skin, and hair of humans (Lian *et al,* 2012). This species is present in abundance on the skin of humans, and is shed directly into coastal waters from bathers. The most well known species in this genus, which is known to cause illness in humans is *S. aureus*. *S. aureus* has a high resistance to salinity, making it a potential threat to other humans using the contaminated water source for recreational purposes. *S. aureus* has also been positively correlated with respiratory, ear, and skin illnesses in humans. While this species is most commonly linked to human illnesses, it is also capable of infecting marine mammals (van Elk *et al,* 2012). The origin of the strain of *S. aureus* that is contracted by marine mammals was most likely from terrestrial sources introduced into the marine environment via runoff (van Elk *et al,* 2012). Studies examining the abundance of *Staphylococcus* over a wet and dry season at a heavily visited coastal area observed increased abundance of *S. aureus* during the wet season (Curiel-Ayala *et al,* 2012). The trends seen in my data also show increased abundance of *Staphylococcus* during Florida's wet season.

*Clostridia*

Within the class *Clostridia* is the genus *Clostridium*. This genus is made up of Gram-positive organisms that form endospores. The formation of endospores allows for this group to survive unfavorable conditions, and proliferate when nutrient rich and favorable conditions occur (Paredes-Saba *et al,* 2011). The four well-known pathogenic species are *Clostridium difficile, C. perfingens, C. tetani,* and *C. botulinum* (www.publichealth.gc.ca). These organisms are known to cause serious illnesses in humans such as botulism, tetanus, colitis, and food poisoning. Pathogenic strains of *Clostridium* are most likely introduced into the marine environment from terrestrial sources. After introduction into marine and coastal waters members of the genus

*Clostridium* have been associated with infectious scleractinian coral diseases. Black band disease is a widespread and highly destructive disease in corals, characterized by a black or red microbial mat band, which migrates from top to bottom in coral colonies. This disease can kill up to 1 cm of healthy coral tissue per day. Multiple species of *Clostridium* have been associated with the occurrence of black band disease in corals (Frias-Lopez, 2002). Yellow band disease in the Caribbean, and White Syndrome in the Indo-Pacific, are two additional coral diseases known to have an association with *Clostridium spp.* Caribbean corals are currently in competition for hard-bottom space with benthic algal species. Many of the algae release primary metabolites as a form of allelopathy. This release of metabolites has been shown to increase microbial growth and may contribute to the disease pathogenicity in the infected corals (Sweet *et al*, 2013).

In this study *Clostridium* was seen in the highest abundance in the winter months of January and February. It is possible that increased abundance of this genus was seen in these months due to increased water temperatures and nutrient availability. As stated previously, temperatures in southern Florida were abnormally high for this time of year and nitrogen and phosphorous ion concentration levels were higher than in other months throughout the year.

*Camplyobacteriales*

*Helicobacter pylori* (*H. pylori)* is one of the most common human pathogens in the world today, infecting roughly 50% of the world's population (Kim *et al,* 2011). It is a Gram-negative bacteria most commonly found in the gastrointestinal system of humans (Vale and Vitor, 2010). Previous studies have shown *H. pylori* is present in many natural water environments including coastal and offshore marine environments (Twing *et al,* 2011; Voytek *et al,* 2005; and Cellini *et al,* 2004), and that it is often more prevalent and in higher abundance in marine waters (Carbone *et al,* 2005 and Cellini *et al,* 2004). Introduction of this species into the coastal marine environment is likely from human fecal contamination, terrestrial runoff, and treated wastewater. Temporal studies completed on *H. pylori* have seen increases in abundance of this organism during the warm summer months (Twing *et al,* 2011). In this study, *Helicobacter spp.* were only

present in month of October and were in very low abundance. The presence of this organism in the month of October could be due to high tidal ranges that occur in this region during this month (Aranda *et al*, 2015 In Press).

*Lactobacillus*

Two pathogenic genera in the Lactobacillus class are present in the Port Everglades Inlet water samples. *Streptococcus spp.* can be pathogenic to humans as well as marine life. *Streptococcus* can be introduced into the environment from humans via fecal contamination. Humans impacted by pathogenic fecal *Streptococcus* have been documented to suffer acute febrile respiratory illness, especially when exposed to increased levels of fecal streptococci (Fleisher *et al,* 1996). *Streptococcus agalactiae* is a zoonotic bacteria and has been shown to cause illness and mortality in humans. Members in this group are most commonly seen to infect bovine animals, fish, and humans causing sepsis, pneumonia, and osteomyelitis. Organisms in this genus have been observed as pathogenic to wild and cultured marine fishes, causing large fish kills in the Gulf of Mexico and the panhandle of Florida up to the Alabama coast. *Streptococcus* was present year-round in this study. The highest abundance of *Streptococcus* was seen in the month of January. Increased freshwater input and warm water conditions could have been the cause of the increased abundance of these organisms. This is also during the prime shipping season in Port Everglades, which may have had an impact on abundance levels of *Streptococcus spp.*

*Enterococcus spp.* are important fecal indicator bacteria, most often utilized to assess fecal contamination on recreational beaches and coastal areas (Aranda *et al*, 2015 In Press; Heaney *et al,* 2014; Wade *et al,* 2003; US Environmental Protection Agency, 1986 and 2004). *Enterococcus spp.* are universally found in feces from humans and animals and are present in almost every environment in the world (Heaney *et al,* 2014). These organisms are able to grow in high salinity conditions, but are known to persist longer in freshwater conditions (Sinton *et al,* 2002). While certain species of enterococci such as *E. faecium* and *E. faecalis* aid in digestion in the gut, certain *Enterococcus spp.* are opportunistic pathogens and cause nosocomial infections such as urinary tract

infections, and abdominal and pelvic infections. Particular strains of enterococci are of concern due to their antibiotic resistance. A recent study examining the number of exceedances of enterococci on recreational beaches in Miami-Dade County, FL from 2000-2010 (Aranda *et al*, 2015 In Press) showed that beaches were only in exceedance of the allowable levels of enterococci 3% of the time. This study examined data generated by the Florida Healthy Beaches Program, which samples weekly. No patterns in regards to rainfall or storms were seen in correlation with enterococci exceedances, although this may be due to the sampling frequency and high decay rate of enterococci in marine waters (Aranda *et al*, 2015 In Press; Sinton *et al,* 2002). The highest levels of eneterococci seen in the Miami-Dade study were in March and October, and this could be due to high tidal levels in October, and possible tourism influences in the month of March overlapping with spring break (Aranda *et al*, 2015 In Press). Interestingly, in a previous study examining presence of pathogens in Port Everglades Inlet, no *Enterococcus spp.* were observed (Campbell *et al,* 2015). This may be due to different sequencing platforms that were used in the studies, or that *Enterococcus spp.* are not prevalent inhabitants in Port Everglades Inlet. More intensive studies would need to be completed to determine presence of *Enterococcus spp.* in Port Everglades Inlet.

*Enterobacteriales*

The Enterobacteriales class was screened for potential pathogens to determine if the coral pathogen, *Serratia marcescens,* was present in inlet water samples. *S. marcescens* is a human fecal bacteria that has been observed as the causative agent of the white pox disease in *Acropora palmata,* an environmentally important coral species in southern Florida. *S. marcescens* was not observed in the top 100 OTUs in the Enterobacteriales class. Possible pathogens seen in this class were *Citrobacter spp., Morganella spp.,* and *Providencia spp.* These organisms are known human pathogens, but have not been documented to cause serious illness to marine organisms or humans after exposure to them in the marine environment. The organism found in highest abundance in this class was *Citrobacter*. Increases in abundance of this organism were seen in the late winter and early spring months. These pathogens were not present in high

abundance at any time throughout the entire year. The presences of Salmonella spp., a common human pathogen, were not detected in the inlet water samples.

*Vibrionales*

Organisms in the Vibrionales class were screened for potential pathogenic bacteria. In the top 100 OTUs, no pathogenic genera were detected in the inlet surface water samples.

*Overall Presence of Pathogens in Port Everglades Inlet*

Port Everglades Inlet surface water samples showed the presence of possible pathogenic genera. Although these organisms were present in the samples, most were not present year-round and were only seen in one or two months throughout the year. These organisms were not in high abundance in any of the samples, with most not being present in the top OTUs in the overall dataset. It is worth noting that when completing sequencing studies, only DNA is available. This being said the viability of these pathogens is unknown. Port Everglades Inlet should be continually monitored for pathogens in the future due to the high influence of anthropogenic factors from tourists, urban development and runoff, and the shipping industry. Future studies examining the presence of pathogens in the port should utilize real-time PCR , DNA sequencing technology, and culturing techniques to obtain counts and viability of cells in the water samples.

*Comparison of Illumina data with 454*

The development of high-throughput DNA sequencing technologies has revolutionized scientist's understanding of microbial community composition in different environments. In recent years the metagenomics community has seen a shift from Roche 454 sequencing technology to Illumin/Solexa platforms, mostly due to the lower cost per base on the Illumina platform (Metzker, 2010; Caporaso *et al.,* 2012). There has been disagreement in the scientific community when determining if the two platforms provide comparable data. In a study completed Luo and colleagues (2012), freshwater samples

with similar complexity to marine samples, but lower than soil samples, were sequenced using both technologies. Their results showed that the two technologies had 90% similarity in assembled contigs and 89% similarity in raw reads and genome abundance for the sample. They concluded that both technologies provide reliable and comparable data for assessing microbial community diversity. Due to the lower cost of sequencing, more accurate contig assembly and less homopolymer read errors, Illumina sequencing may be the better platform for metagenomic analyses (Luo *et al,* 2012;Mardis, 2008; Metzker, 2010).

In this study, the Illumina MiSeq platform was used to sequence Port Everglades Inlet water sample. In a previously published paper in our lab, Port Everglades Inlet water samples were sequenced using the Roche 454 pyrosequencing technology. Results for both studies were analogous, with the major bacterial taxa and their relative abundances showing homogeneity across both studies. These data indicate that both platforms produce similar and reliable results when comparing data generated from the same sampling site. With the costs of Illumina being roughly a quarter of those produced in 454 pyrosequencing, it can be stated that Illumina sequencing is the better option for cost-effective sequencing projects. Regardless of cost, both platforms provide reliable and relevant data.

**Section VI: Importance and Significance**

High-throughput sequencing technologies are of utmost importance in marine microbial community studies due to the limitations of VBNC microbes in marine environments (Handelsman, 2004; Giovannoni *et al,* 1990). Previous to the use of high-throughput DNA sequencing, culture-based and PCR approaches encompassed majority of the information known about marine microbes. Due to the unculturable nature of the marine microbes, scientists were left with a grossly underestimated knowledge of true diversity of marine microbial communities. As DNA sequencing technologies continue to develop increased read lengths and more reads per sample, true microbial composition of a community can be determined. This will be important for future monitoring efforts in marine ecosystems, as well as determining rare, but important environmental indicator organisms that formerly went undetected in microbial studies.

Port Everglades Inlet is a major seaport in southern Florida, and is a major connection into the Atlantic Ocean. Due to the busy nature of the port, it is constantly under anthropogenic influences from humans, terrestrial runoff, urban development, and the shipping industry (cargo and cruise ships, and private boats). The inlet waters flow out into the Atlantic Ocean and directly offshore onto a coral reef and adjacent beaches, making it a point source of pollution to these environments (Banks *et al,* 2008; Carrie-Futch *et al,* 2011; Stamates *et al,* 2013). Previous to this study, majority of the research and monitoring completed in the inlet used culture-based and qPCR approaches, and focused on presence of fecal indicator bacteria. Only one other study has utilized high-throughput DNA sequencing technology to generate baseline knowledge of the microbial community composition in Port Everglades Inlet, but sampling scale was much larger, obtaining samples quarterly as opposed to weekly. Small scale weekly sampling allowed for fluctuations in community composition to be determined on a monthly and seasonal basis over an entire year. This information is important for future public health and environmental monitoring efforts in the port and will provide information regarding natural fluctuations of the microbial community as well as fluctuations caused by pollution and other anthropogenic influences.

## Section VII: Conclusions

The goal of this study was to characterize the marine microbial community composition of the Port Everglades Inlet surface water microbiome and it's fluctuations throughout a 1-year timespan. This is the first study, to my knowledge, that utilized Illumina sequencing technology and that sampled on a weekly basis to analyze microbes in Port Everglades Inlet. Results indicated that there were significant differences in alpha diversity, especially when comparing the microbial community in the month of August, with the months of December, October, and November, and this is most likely due to the low community richness and abundance seen in the month of August, which also had significantly low precipitation levels. Significant differences in beta diversity were seen when comparing months and seasons but no significant differences were observed for location. Species richness and diversity varied by month and by season, but was

dependent on individual bacterial organisms. No drastic increases in species richness or diversity were seen in Florida's wet season, which was hypothesized. Environmental metadata such as rainfall, salinity, water temperature, and nutrient concentration had impacts on the microbial community composition of Port Everglades Inlet waters, specifically in regards to months in the dry season. The dry season in Florida for the 2013-2014 year, was significantly warmer than has been seen in previous years. In addition to the warm temperatures, increases in nitrogen and phosphorous concentration were also observed in these months, possibly creating optimal conditions for bacterial growth during these times. Increased microbial abundance and richness were seen in the early spring and late summer months, most likely due to increased temperatures, UV radiation, and precipitation, which was expected.  Pathogenic genera were detected in the inlet but were not consistently present throughout the year and were at low abundances. In the months that the pathogens were detected it is possible that they were present due to the prime shipping season in Port Everglades Inlet which runs from November to April and also correlated with tourist season in south Florida.

In comparison to the previous study by Campbell et al (2015) completed on Port Everglades Inlet and surrounding waters, the overall community compositions were comparable. The microbes present in the inlet waters were also comparable to other marine microbial communities outlined in different marine and coastal environments. The utilization of Illumina sequencing allowed for a much higher numbers of reads per sample than were seen in the previous inlet study, but also did not have as long of a read length seen with 454-pyrosequencing technology. Both studies on the Port Everglades Inlet microbiome yielded complementary data creating a strong baseline of the microbiome present in the inlet. These studies can be utilized by the county and public health officials, who complete routine monitoring on port waters, as well as by environmental scientists looking to see what the impacts of the microbial community in Port Everglades Inlet might have on the surrounding coastal beaches and adjacent coral reef.

Future microbial community composition studies in Port Everglades Inlet should utilize multiple molecular techniques in order to determine viability, abundance, and

ecological roles these organisms play in their environment. The use of qPCR and culture-based techniques paired with DNA sequencing would provide in-depth information regarding viable pathogens present in Port Everglades Inlet, and could be of use in determining if the Inlet is a source of disease for the offshore coral reef ecosystem, as well as if the beaches adjacent to the inlet are drastically impacted by the expelled inlet waters. A multi-omic study completed in the inlet would supplement the current study by providing information on the ecological roles the microbes play in this particular niche. It would be interesting to see if roles of the microbes in the inlet differ from roles played in other marine ecosystems.

**References:**

Ahrendt, S.R., Mobberley, J.M., Visscher, P.T., Koss, L.L., and Foster, J.S. (2014) Effects of elevated carbon dioxide and salinity on the microbial diversity in lithifying microbial mats. *Minerals*, **4**: 145-169.

Alves, L., Siniscalchi, B., Campante Vale, I., Dell'Isola, J., Chernicharo, C.A., and Araujo, J.C. (2015) Enrichment and activity of methanotrophic microorganisms from municipal wastewater sludge. *Environmental Technology*, **36(12)**: 1563-1575

Ansorge, W. (2009). Next-generation DNA sequencing techniques. *New Biotechnology*. ***25*(4)**: 195-203.

Aranda, D., Lopez, J.V., Solo-Gabriele, H.M., Fleisher, J.M. (2015). Using probabilities of enterococci exceedance and logistic regression to evaluate long term weekly beach monitoring data. *Journal of Water and Health.* In press.

Arrigo, K.R. (2005). Marine microorganisms and global nutrient cycles. *Nature*. ***437*(15)**: 349-355.

Australian Institute of Marine Science.(2013) Marine microbes. *Australian Government.* Retrieved on 8 October, 2013. http://www.aims.gov.au/docs/research/marine-microbes/microbes/microbes.html

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L., Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*. 10, 257-263

Banks, K.E., Riegl, B.M., Richards, V.P., Walker, B.E., Helmle, K.P., Jordan, L.K.B., Phipps, J., Shivji, M., Spieler, R.E., and Dodge, R.E. (2008) Chapter 5: The reef tract of continental Southeast Florida (Miami-Dade, Broward, and Palm Beach Counties, USA). In: Riegl, B., Dodge, R.E. (eds) Coral Reefs of the USA. Springer-Verlag, Dordrecht, 175-220.

Birtel, J., Walser, JC, Pichon, S., Burgmann, H., and Matthews, B.(2015) Estimation bacterial diversity for ecological studies: Methods, metrics, and assumptions. *PLoS ONE* **10.** DOI: 10.1371/journal.pone.0125356 . 2015.

Brandes, J.A., Devol, A.H., Deutsch, C. (2007) New developments in the marine nitrogen cycle. *Chem. Rev.* ***107***: 577-589.

Bretter, I., Christen, R., and Hofle, M. (2003) *Idiomarina baltica* sp. Nov., a marine bacterium with a high optimum growth temperature isolated from surface water of the

central Baltic Sea. *International Journal of Systematic and Evolutionary Microbiology*. **53**: 407-413. DOI 10.1099/ijs.0.02399-0

Bridoux, M.C., Neibauer, J., Ingalls, A.E., Nunn, B.L., Keil, R.G. (2015) Suspended marine particulate proteins in coastal and oligotrophic waters. *Journal of Marine Systems*, **143**: 39-48. doi:10.1016/j.jmarsys.2014.10.014

Brown, M.V., Lauro, F.M., Demaere, M.Z., Muir, L., Wilkins, D. et al. (2012) Global biogeography of SAR11 marine bacteria. *Molecular Systems Biology*, **8:595**

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., ... & Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science*, ***273*(5278)**: 1058-1073.

Campbell, A.M., Fleisher, J., Sinigalliano, C., White, J.R., and Lopez, J.V. (2015) Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of southeast Florida. *MicrobiologyOpen*. 1-19. doi: 10.1002/mbo3.245

Campbell, B.J., and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *The ISME Journal*. **7**: 210-220.

Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010). The evolution and future of Earth's nitrogen cycle. *Science*. ***330***: 192-196.

Caporaso, J.G., Kuczynski, Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequence data. *Nature Methods*. ***7*(5)**: 335-336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *International Society for Microbial Ecology*. ***6***: 1621-1624.

Carbone, M., Maugeri, T.L., Gugliandolo, C., La Camera, E., Biondo, C., and Fera, M.T. (2005) Occurrence of *Helicobacter pylori* DNA in the coastal environment of southern Italy (Straits of Messina). *Journal of Applied Microbiology,* **98(3)**: 768-774.

Carrie Futch, J., Griffin, D.W., Banks, K., and Lipp, E.K. (2011). Evaluation of sewage source and fate on southeast Florida coastal reefs. *Marine Pollution Bulletin*. ***62***: 2308-2316.

Carsey, T.P., Stamates, S.J., Amornthammarong, N., Bishop, J.R., Bloetscher, F., Brown, C.J., Craynock, J.F., Cummings, S.R., Paul Dammann, W., Davis, J., Featherstone, C.M., Fischer, C.J., Goodwin, K.D., Meeroff, D.E., Proni, J.R., Sinigalliano, C.D., Swart, P.K., and Jia-Zhong Zhang. NOAA Data Report AOML-XX Boynton Inlet 48-hour Sampling Intensives, June and September, 2007.

Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W., and Waterbury, J. B. (1981). Prokaryotic cells in the hydrothermal vent tube worm Riftia pachyptila Jones: possible chemoautotrophic symbionts. *Science*, ***213*****(4505)**: 340-342.

Cellini, L., Del-Vecchio, A., Di Candia, M., Di Campli, E., Favaro, and M., and Donneli, G. (2004) Detction of free and plankton-associated *Helicobacter pylori* in seawater. *Journal of Applied Microbiology*, **97(2)**: 3881-3884.

Cephas, K.D., Kim, J., Mathai, R.A., Barry, K.A., Dowd, S.E., Meline, B.S., and Swanson, K.S. (2011). Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. *PLoS One*. ***6*****(8)**: N/A.

Chial, H. (2008). DNA sequencing technologies key to the human genome project. *Nature Education***. 1.**

Codispoti, L.A., Brandes, J.A. Christensen, J.P., Devol, A.H. Naqvi, S.W.A., Paerl, H.W., and Yoshinari, T. (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci. Mar. 65*(2). 85-105.
Corliss, J.B., Dymond, J., Gordon, L.I., Edmond, J.M., von Herzen, R.P., Ballard, R.D., … and van Andel, T.H. (1979). Submarine thermal springs on the Galapagos Rift. *Science*. ***203*****(4385)**: 1073-1083.

Curiel-Ayala, F., Quinones-Ramirez, EI., Pless, RC., and Gonzalez-Jasso, E. (2012) Comparative studies on *Enterococcus, Clostridium perfringens,* and *Staphylococcus aureus* as quality indicators in tropical seawater at a Pacific Mexican beach resort. *Marine Pollution Bulletin,* **64(10)**: 2193-2198. doi:10.1016/j.marpolbul.2012.07.052

Dang, H., Li, T., Chen, M., and Huang, G. (2008) Cross-Ocean Distribution of *Rhodobacerales* Bactera as Primary Surface Colonizers in Temperate Coastal Marine Waters. *Applied and Environmental Microbiology,* **79** **(1)**: 52-60. doi:10.1128/AEM.01400-07

Dang, H.Y., and Lovell, C.R. (2000) Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. *Applied Environmental Microbiology*, **66**: 467–475

Dang, H., Chen, R., Wang, L., Shao, S., Dai, L., Ye, Y., Guo, L., Huang, G., and Klotz, M.G. (2011). Molecular characterization of putative biocorroding microbiota with a novel niche detection of *Epsilon-* and *Zetaproteobacteria* in Pacific Ocean coastal seawaters. *Environmental Microbiology*, **13(11):** 3059-3074. doi:10.1111/j.1462-2920.2011.02583.x

Dave, M., Higgins, P.D., Middha, S., Rioux, K.P. (2012). The human gut microbiome: Current knowledge, challenges, and future directions. *Translational Research.* ***160*****(4):** 246-257.

DeLong, E.F. and Karl, D.M. (2005). Genomic perspectives in microbial oceanography. *Nature*. ***437*****:** 336-342.

DeLong, E.F. (2009). The microbial ocean from genomes to biomass. *Nature*. ***459*****(7244):** 200-206.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., and Hess, W. R. (2003). Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences*, ***100*****(17):** 10020-10025.

Dyhrman, S.T., Ammerman, J.W., and Van Mooy, B.A.S. (2007). Microbes and the marine phosphorus cycle. *Oceanography*. ***20*****(2):** 110-116.

Falksowski, P.G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of $CO_2$ in the ocean. *Nature*. ***387*****:** 272-275.

Felbeck, H. (1981). Chemoautotrphic potential of the hydrothermal vent tube worm, Riftia pachyptila Jones Vestimentifera. *Science.* ***213*****(4505):** 336-338.

Finkl, C.W. and Charlier, R.H. (2003). Sustainability of subtropical coastal zones in southeastern Florida: Challenges for urbanized coastal environments threatened by development, pollution, water supply, and storm hazards. *Journal of Coastal Research*. ***19*****(4):** 934-943.

Fleisher, J.M., Kay, D., Salmon, R.L., Jones, F., Wyer, M.,and Godfree, A.F. (1996). Marine waters contaminated with domestic sewage: Nonenteric illnesses associated with bather exposure in the United Kingdom. American Journal of Public Health**, 86(9):** 1228-1234.

Flombaum,P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K.W, Lomas, M.W., Tully, B.J., Sachdeva, R., Heidlberg, K.B., and Heidelberg, J.F. (2014) Comparative genomics of planktonic *Flavobacteriaceae* from the

Gulf of Maine using metagenomic data. *Microbiome,* **2(34).**

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jioa, N., Karl, D.M., Li, W.K.W., Lomas, M.W., Veneziano, D., Vera, C.A., Vrugt, J.A., Martiny, A.C. (2013). Present and future global distributions on marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *PNAS*. ***110*****(24):** 9824-9829. www.pnas.org/cgi/doi/10.1073/pnas.1307701110.

Froelich, P.N., Bender, M.L., and Luedtke, N.A. (1982). The marine phosphorus cycle. *American Journal of Science.* ***282*****:** 474-511.

Frias-Lopez, J., Zerkle, A.L., Bonheyo, G.T., and Fouke, B. (2002). Partitioning of bacterial communities between seawater and healthy, black band diseases, and dead coral surfaces. *Applied and Environmental Microbiology,* **68(5)**: 2214-2228. DOI: 10.1128/AEM.68.5.2214–2228.2002

Freitas, S., Hatosy, S., Fuhrman, J.A., Huse, S.M., Welch, D.B.M., Sogin, M.L., and Martiny, A.C. (2012). Global distribution and diversity of marine *Verrucomicrobia*. *The Isme Journal*. ***6*****:** 1499-1505.

Gaill, F. (1993). Aspects of life of development at deep sea hydrothermal vents. *The FASEB Journal***,** ***7*****(6):** 558-565.

Georges, A.A., El-Swais, H., Craig, S., Li, K.W.W, and Walsh, D,A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME*, **8:** 1301-1313.

Ghiglione, J.F., and Murray, A.E. (2012) Pronounced summer to winter differences and higher wintertime richness in coastal Antarctic marine bacterioplankton. *Environmental Microbiology,* **14(3):** 617-629. doi :10.1111/j.1462-2920.2011.02601.x

Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., Somerfield, P.J., and Muhling, M. (2010). The taxonomic and functional diversity of microbes at temperate coastal site: A ‘multi-omic’ study of seasonal and diel temporal variation. *PloS One.* ***5*****(11).**

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton.

Glockner, F.O., Fuchs, B.M., Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Applied and Environmental Micorbiology*. ***65*****(8):** 3721-3726.

Gomez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E., and Amann, R. (2010). Distinct flavobacterial communities in contrasting water masses of the North Atlantic. *ISME*, **4(4):** 472-487. :10.1111/j.1462-2920.2011.02601.x

Gonzalez, J.M., and Moran, M.A.(1997). Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. *Applied and Environmental Microbiology,* **63(11):** 4237-4242.

Gruber, N. (2004). Dynamics of the marine nitrogen cycle and its influence on atmospheric CO2 variations. In *The ocean carbon cycle and climate*. 97-148. Springer, Netherlands.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*. ***68*(4):** 669-685.

Herbert, R.A. (1999). Nitrogen cycling in coastal marine ecosystems. *FEMS Microbiology Reviews*. ***23:*** 563-590.

Hobbie, J.E., Daley, J. Jasper, S. (1977). Use of nucleopore filters for counting bacteria by fluorescence microscopy. *Appl. Environ. Microbiol.* ***33*(5):** 1225-1228.

Huang, J.Q., Dang, H.Y., Li, T.G., Huang, G.Q., and Pan, L.Q. (2008) Research on diversity and community succession of marine attachment bacteria. *Chin High Tech Lett,* **18:** 971– 978.

Huggett, M.J., Hayakawa, D.H., and Rappe, M.S. (2012). Genome sequence of strain HIMB624, a cultured representative from the OM43 clade of marine *Betaproteobacteria*. *Standards in Genomic Sciences,* **6(1):** 11-20. doi: 10.4056/sigs.2305090

Jackson, C.R., and Vallaire, S.C. (2009) Effects of salinity and nutrients on microbial assemblages in Louisiana wetland sediments. *Wetlands,* ***29*(1):** 277-287.

Janda, J.M., and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinincal Microbiology,* 2761-2764.

Jenson, S., Duperron, S., Birkeland, NK, and Hovland, M. (2010). Intracellular *Oceanospirillales* bacteria inhabit gills of *Acesta* bivalves. *FEMS Microbiology Ecology,* DOI: http://dx.doi.org/10.1111/j.1574-6941.2010.00981.x 523-533.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Malcom, E., Woodward, S., Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*. ***311*:** 1737-1740.

Jones, P.R., Cottrell, M.T., Kirchman, D.L., and Dexter, S.C. (2007) Bacterial community structure of biofilms on artificial surfaces in an estuary. *Microbial Ecology*, **53:** 153–162.

Kelley, S.T. and Gilbert, J.A. (2013). Studying the microbiology of the indoor environment. *Genome Biology*. ***14*(2):** 202.

Kim, S.S., Ruiz, V.E., Carroll, J.D., and Moss, S.F. (2011). *Helicobacter pylori* in the pathogenesis of gastric cancer and gastric lymphoma. *Cancer Letters*, **305(2):** 228-238. DOI: http://dx.doi.org/10.1016/j.canlet.2010.07.014

Koleff,P., Gaston, K.J., and Lennin, J.J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology.* ***72:*** 367-382.

Kong, H.H. (2011). Skin microbiome: Genomics-based insights into the diversity and role of skin microbes. *Trends in Molecular Medicine*. ***17*(6):** 320-328.

Kuczynski, J., Stombaugh, J., Walters, W.A., Gonzalez, A., Caporaso, J.G., and Knight, R. (2011). Using QIIME to analyze 16s rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics*. 10-7.

Krause, E., Wichels, A., Gimenez, L., Lunau, M., Schillhabel, M.B., and Gerdts, G. (2012). Small changes in pH have direct effects on marine bacterial community composition: A microcosm approach. *PLoSONE* DOI: 10.1371/journal.pone.0047035

Kris, A.E. (1963). Marine Microbiology. Tran. J.M., Shewan and Z. Kabata. Oliver and Boyd, London.

Kutschera U., and Schauer, S., (2012) Prokaryotic biodiversity in marine versus terrestrial ecosystems: Methylobacteria and research ethics. Journal of Marine Science: *Research and Development*, **2.**

Lefort, T., and Gasol, J.M. (2013) Global-scale distributions of marine surface bacterioplankton groups along gradients of salinity, temperature, and chlorophyll: a meta-analysis of fluorescence *in situ* hybridization studies. *Aquatic Microbial Ecology.* **70**: 111-130. doi: 10.3354/ame01643

Lian, P.Y., Maseko, T., and Rhee, M., N.G K. (2012) The antimicrobial effects of cranberry against *Staphylococcus aureus*. *Food Science and Technology International*, **18:** 179-186.

Lin, Y.C., Campbell, T., Chung, C.C., Gong, G.C., Chiang, K.P., and Worden, A.Z. (2012). Distribution patterns and phylogeny of marine stramenopiles in the North Pacific

Ocean. *Applied Environmental Microbiolgy,* **78(9):** 3387-3399. doi:10.1128/AEM.06952-11

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K.T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE*, **7(2):** e30087.

Mallin, M.A., Williams, K.E., Cartier Esham, E., and Lowe, R.P. (2000). Effect of human development on bacteriological water quality in coastal watersheds. *Ecological Applications*. ***10*(4):** 1047-1056.

Marcus, N. (2004). An overview of the impacts of eutrophication and chemical pollutants on copepods of the coastal zone. *Zoological Studies*. ***43*(2):** 211-217.

Mardis, E. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. ***24*(3):** 133-141.

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature*. **470:** 198-203.

Maxam, A.M., Gilbert, W. (1977). A new method for sequencing DNA. *Proc. National Academy of Sciences. USA*. ***74*(2):** 560-564.

McIntyre, A.D. (1990). Sewage in the sea, Annex XII of *State of the Marine Environment.* GESAMP Reports and Studies. No. 39.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE,* **8(4):**e61217.

Meron, D., Rofoldo-Metalpa, R., Cunning , R., Baker, A.C., Fine, M., and Banin, E. (2012). Changes in coral microbial communities in response to a natural pH gradient. I*SME*, **6**: 1775-1785. doi:10.1038/ismej.2012.19;

Metzker, M. (2010). Sequencing technologies- the next generation. *Nature Reviews*. **11:** 31-46.

Meyer-Reil, L.A. and Koster, M. (2000). Eutrophication of marine waters: Effects on benthic microbial communities. *Marine Pollution Bulletin*. ***41*:** 255-263

Mitsui, H., Gorlach, K., Lee, H., Hattori, R., and Hattori, T. (1997) Incubation time and media requirements of culturable bacteria from different phylogenetic groups. *Journal of Microbiological Methods*, **30:** 103-110.

Mulheron, B. (2015) Microbial community assembly found with sponge orange band disease in *Xestospongia muta* (Giant orange barrel sponge). Master's Thesis Unpublished. Nova Southeastern University Oceanographic Center.

Needham, D.M., Chow, C.T., Cram, J.A., Sachdeva, R., and Parada, A., and Fuhrman, J. (2013). Short-term observations of marine bacterial and viral communities; Patterns, connections and resilience. *ISME***, (7**): 1274-1285.

Neufeld, J.D., Boden, R., Moussard, H., Schafer, H., and Murrell, JC. (2008) Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Applied and Environmental Microbiology*, **74(23):** 7321-7328. doi:10.1128/AEM.01266-08

NOS (National Ocean Service, National Oceanic and Atmospheric Administration), 2011. Coast Pilot 4. Atlantic Coast Cape Henry to Key West, 43rd ed., Chapter 10, St. John's River to Miami, pp. 328-332.

Paerl, H.W., Dyble, J., Moisander, P.H., Noble, R., Piehler, M.F., Pinckney, J.L., Steppe, T.F., Twomey, L., and Valdes, L.M. (2006). Microbial indicators of aquatic ecosystem change: Current applications to eutrophication studies. *FEMS Microbiology Ecology*. ***46*(3):** 233-246.

Pagani, I., Liolios, K., Jansson, J., Chen, I.A., Smirnova, T., Nosrat, B., Markowitz, V.M., and Kyrpides, N.C. (2012) The Genomes OnLine Database(GOLD) v.4: status of genomic and metagenomic projects and their associated metdata. *Nucl. Acids Res, **40:*** 571-579.

Paredes-Sabja, D., Setlow, P., and Sarker, M.R. (2011). Germination of spores of *Bacillales* and *Clostridiales* species: mechanisms and proteins involved. *Trends in Microbiology*. **19(2)**: 85-94.

Partensky, F., Blanchot, J., and Vaulot, D. (1999). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bulletin de l'Institut oceanographique*.

Pedros-Alio, C. (2006). Genomics and marine microbial ecology. *International Microbiology. **9:*** 191-197.

Pomeroy, L.R. (1974). The ocean's food web, a changing paradigm. *BioScience. **24***(**9**)**:** 499-504. ***9:*** 457-475.

N/A. (n.d.) 454 Overview Tutorial: de novo OTU picking and diversity analyses using 454 data. *Quantitative Insights into Microbial Ecology*. Date retrieved: 1/7/2014. http://qiime.org/tutorials/tutorial.html

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome centre's improvements to the Illumina sequencing system. *Nature Methods*. ***5*(12):** 1008-1010.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods.* ***6*****(9):** 639-641.

Reddy, T.B.K., Thomas, A., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, .J, Pagani, I., Lobos, E., and Kyrpides, N. (2014) The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research*, *doi: 10.1093/nar/gku950*

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., and Chisholm, S. W. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*, ***424*****(6952):** 1042-1047.

Rohwer, F., & Youle, M. (2010). *Coral reefs in the microbial seas*. United States: Plaid Press.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, ***239*****(4839):** 487-491.

SAS Institute
Cary, NC 27513-2414, USA

Schlesner, H., Jenkins,C., and Staley, J.T. (2006) The phylum Verrucomicrobia: A phylogenetically heterogeneous bacterial group. *Prokaryotes*. ***7:*** 881-896.

Schuster, S.C, (2008). Next-generation sequencing transforms today's biology. *Nature Methods.* ***5*****(1):** 16-18.

Schwenn, J.D., Depka, B., and Hennies, H.H. (1976) Assimilatory sulfate reduction in chloroplasts: Evidence for the participation of both stromal and membrane-bound enzymes. Plant and Cell Physiology, **17**: 165-176.

Seki, H. (1972). The role of microorganisms in the marine food chain with reference to organic aggregates. *Mem. Ist. Ital. Idrobiol.* ***28*****:** 245-259.

Shendure, J., Ji H. (2008). Next generation DNA sequencing. *Nature Biotechnology.* ***26(*****10):** 1135-1145.

Snieszko, S.F. (1974). The effects of environmental stress on outbreaks of infectious diseases of fishes. *Journal of Fish Biology*. ***6*****:** 197-208.

Somboonna, N., Assawamakin, A., Wilantho, A., Tangphatsornruang, S., and Tongsima, S. (2012). Metagenomic profiles of free-living archaea, bacteria and small eukaryotes in coastal areas of Sichang island, Thailand. *BMC Genomics*. ***13*****(7):** 2-19.

Stahl, D. A., Lane, D. J., Olsen, G. J., and Pace, N. R. (1984). Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science*, ***224*****(4647):** 409-411.

Stamates, S.J., Bishop, J.R., Carsey, T.P., Featherstone, C.M., Gidley, M.L., Kelble, C.R., Kotkowski, R.M., Roddy, R.J. (2013). Port Everglades Flow Measurement System. *NOAA Technical Report, OAR-AOML-42*. 1-22

Sutherland, K.P., Shaban, S., Joyner, J.L., Porter, J.W., and Lipp, E.K. (2011). Human pathogen shown to cause disease in the threatened elkhorn coral *Acorpora palmata. PloS One*. ***6*****(8):** N/A.

Sweet, M.J., Bythell, J.C., and Nugues, M.M. (2013). Algae ad reservoirs for coral pathogens. *PLOS ONE,* DOI: 10.1371/journal.pone.0069717

Thaler, M., and Lovejoy, C. (2014) Enviromental selection of marine stramenophile clades in the Arctic Ocean and coastal waters. *Polar Biology*, **37(3):** 347-357.

Twing, K.I., Kirchman, D.L., and Campbell, B.J. (2011). Temporal study of *Helicobacter pylori* presence in coastal freshwater, estuarine, and marine waters. *Water Research*, **45:** 1897-1905.

US Environmental Protection Agency. (1986). Ambient water quality criteria for bacteria 1986. EPA 440/5-84-002. Office of Water Regulations and Standards, US Envrionmental Protection Agency, Washington, DC.

*US Environmental Protection Agency. 2004. Implementation guidance for ambient water quality criteria for bacteria. EPA-823-B-04-002. US Environmental Protection Agency, Washington, DC.*

Vale, F.F., and Vitor, J.M.B. (2010). Transmission pathway of *Helicobacter pylori*: Does food play a role in rural and urban areas? *International Journal of Food Microbiology,* **138**: 1-12.

Veneziano, D., Vera, C.S., Vrugt, J.A., and Martiny, J.C. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *PNAS*, **110(24):** 9824-9829doi:10.1073/pnas.1307701110

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., ... and Beasley, E. (2001). The sequence of the human genome. *Science*, ***291*****(5507):** 1304-1351.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern,A.L., Rusch, D., Eisen, J.A., Wu, D., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* **304(5667):** 66-74.

Voytek, M.A., Ashen, J.B., Fogarty, L.R., Kirshtein, J.D., Landa, E.R. (2005). Detection of *Helicobacter pylori* and fecal indicator bacteria in five North American rivers. *Journal of Water and Health,* **3(4)**: 405-422.

Wade, T.J., Pai, N., Eisenberg, J.N., Colford, J.M. (2003). Do US Environmental Protection Agency water quality guidelines for recreational water prevent gastrointestinal illness? A systematic review and meta-analysis. *Environmental Health Perspective*, **111:** 1102-1109.

Walker, B.K., Gilliam, D.S., and Dodge, R.W., 2008. Mapping coral reef habitats in southeast Florida using a combined technique approach. *J. Coastal Res*. **24:** 1138-1150.

Walker, B.K. 2012. Spatial analyses on benthic habitats to defines coral reef habitats to define coral reef ecosystem regions and potential biogeographic boundaries along a latitudinal gradient. *PLoS One.* 7:e30466

Whitfield, J. (2003). First ocean bacteria sequenced. *Nature.* ***10***: 6.

Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon.* ***21*****(2/3):** 213-251.

Wiebe, W.J. and Pomeroy, L.R. (1972). Microorganisms and their association with aggregates and detritus in the sea: A microscopic study. *Mem. Ist. Ital. Idrobiol.* ***29****:* 325-352.

Windom, H.L. (1992). Contamination of the marine environment from land-based sources. *Marine Pollution Bulletin.* ***25***: 32-36.

Xu, H. S., Roberts, N., Singleton, F. L., Attwell, R. W., Grimes, D. J., and Colwell, R. R. (1982). Survival and viability of nonculturableEscherichia coli andVibrio cholerae in the estuarine and marine environment. *Microbial Ecology*, ***8*****(4):** 313-323.

APPENDICES

APPENDIX 1: Table of sample IDs, location, number of reads per sample, month sample taken and the date sample was taken.

| Sample Name | Number of Reads | | Date Sampled |
|---|---|---|---|
| **PE12 M9** | 11512 | | 71013 |
| **PE15 M9** | 14154 | July | 71913 |
| **PE17 BB** | 11346 | | 72313 |
| **PE18 M9** | 15161 | | 72313 |
| **PE20 BB** | 9507 | | 80313 |
| **PE21 M9** | 9418 | August | 80313 |
| **PE23 BB** | 9500 | | 81013 |
| **PE26 BB** | 17611 | | 81713 |
| **PE27 M9** | 25521 | | 81713 |
| **PE29 BB** | 17586 | | 82013 |
| **PE30 M9** | 19133 | | 82013 |
| **PE32 BB** | 14880 | | 90513 |
| **PE33 M9** | 19612 | September | 90513 |
| **PE35 BB** | 23523 | | 91213 |
| **PE36 M9** | 25149 | | 91213 |
| **PE38 BB** | 10177 | | 92013 |
| **PE39 M9** | 6914 | | 92013 |
| **PE41 BB** | 15134 | | 92713 |
| **PE42 M9** | 13586 | | 92713 |
| **PE44 BB** | 14079 | | 100313 |
| **PE45 M9** | 16402 | | 100313 |
| **PE47 BB** | 11118 | October | 101113 |
| **PE48 M9** | 11936 | | 101113 |
| **PE50 BB** | 18833 | | 101713 |
| **PE51 M9** | 18211 | | 101713 |
| **PE52 BB** | 15871 | | 102613 |
| **PE55 BB** | 14755 | | 110113 |
| **PE56 M9** | 11926 | | 110113 |
| **PE61 BB** | 32354 | | 111413 |
| **PE62 M9** | 13163 | | 111413 |
| **PE64 M9** | 12391 | November | 112113 |
| **PE66 BB** | 12748 | | 112113 |
| **PE67 BB** | 8859 | | 120513 |
| **PE68 M9** | 6066 | December | 120513 |

| **PE71 M9** | 14807 | | 121213 |
|---|---|---|---|
| **PE73 BB** | 7975 | | 121513 |
| **PE74 M9** | 6283 | | 121513 |
| **PE79 BB** | 22560 | | 11014 |
| **PE82 BB** | 19577 | January | 11514 |
| **PE83 M9** | 17723 | | 11514 |
| **PE85 BB** | 29566 | | 11514 |
| **PE86 BB** | 40611 | | 12514 |
| **PE87 M9** | 15039 | | 12514 |
| **PE89 BB** | 17367 | | 13114 |
| **PE90 M9** | 20727 | | 13114 |
| **PE92 BB** | 7861 | February | 20714 |
| **PE93 M9** | 16486 | | 20714 |
| **PE95 M9** | 31575 | | 21214 |
| **PE96 M9** | 28479 | | 21214 |
| **PE98 BB** | 26941 | | 22214 |
| **PE99 M9** | 37174 | | 22214 |
| **PE101 BB** | 11360 | | 22814 |
| **PE102 M9** | 18870 | | 22814 |
| **PE104 BB** | 24331 | March | 30814 |
| **PE105 M9** | 19252 | | 30814 |
| **PE107 BB** | 26295 | | 31414 |
| **PE108 M9** | 25020 | | 31414 |
| **PE110 BB** | 14573 | | 32114 |
| **PE111 M9** | 16854 | | 32114 |
| **PE113 BB** | 16702 | | 32814 |
| **PE114 M9** | 27858 | | 32814 |
| **PE119 BB** | 13035 | | 41114 |
| **PE120 M9** | 17671 | | 41114 |
| **PE122 BB** | 80122 | April | 41814 |
| **PE123 M9** | 5666 | | 41814 |
| **PE125 BB** | 12517 | | 42514 |
| **PE126 M9** | 15088 | | 42514 |
| **PE128 BB** | 12832 | | 50214 |
| **PE129 M9** | 12113 | May | 50214 |
| **PE131 BB** | 8883 | | 50714 |
| **PE132 M9** | 14352 | | 50714 |
| **PE134 M9** | 13711 | | 51714 |
| **PE136 BB** | 11179 | | 51714 |

| | | | |
|---|---|---|---|
| **PE138 M9** | 16157 | | 52314 |
| **PE140 BB** | 14518 | | 53014 |
| **PE141 M9** | 18933 | | 53014 |
| **PE143 BB** | 24207 | June | 60614 |
| **PE144 M9** | 19952 | | 61314 |
| **PE146 BB** | 14031 | | 61314 |
| **PE147 M9** | 16674 | | 61314 |
| **PE149 BB** | 11147 | | 62114 |
| **PE150 M9** | 18294 | | 62114 |

APPENDIX 2: Table of sample IDs with raw environmental metadata.

| Sample Name | Chloride Ion (mol/L) | Sulfate Ion (mol/L) | Phosphate Ion (mol/L) | Salinity | Water Temperature (°C) | Rainfall (inches) |
|---|---|---|---|---|---|---|
| **PE12 M9** | 0.519 | 0.0234 | Below Cal | 26.0 | 30.0 | 12.71 |
| **PE15 M9** | 0.218 | 0.0111 | Below Cal | 14.5 | 30.0 | 12.71 |
| **PE17 BB** | 0.337 | 0.0162 | Below Cal | 17.0 | 30.0 | 12.71 |
| **PE18 M9** | 0.354 | 0.017 | Below Cal | 17.0 | 30.0 | 12.71 |
| **PE20 BB** | Below Cal | Below Cal | Below Cal | 22.5 | 30.0 | 4.43 |
| **PE21 M9** | Below Cal | Below Cal | Below Cal | 22.5 | 30.0 | 4.43 |
| **PE23 BB** | Below Cal | Below Cal | Below Cal | 23.0 | 30.0 | 4.43 |
| **PE26 BB** | 0.461 | 0.0215 | Below Cal | 22.5 | 29.4 | 4.43 |
| **PE27 M9** | 0.404 | 0.0196 | Below Cal | 22.5 | 29.4 | 4.43 |
| **PE29 BB** | 0.538 | 0.0243 | Below Cal | 29.0 | 28.9 | 4.43 |
| **PE30 M9** | 0.561 | 0.0254 | Below Cal | 29.0 | 28.9 | 4.43 |
| **PE32 BB** | 0.435 | 0.0205 | Below Cal | 24.0 | 28.9 | 10.48 |
| **PE33 M9** | 0.437 | 0.02 | Below Cal | 24.0 | 28.9 | 10.48 |
| **PE35 BB** | 0.401 | 0.019 | Below Cal | 24.0 | 28.3 | 10.48 |
| **PE36 M9** | 0.405 | 0.0234 | Below Cal | 24.0 | 28.3 | 10.48 |
| **PE38 BB** | Below Cal | Below Cal | Below Cal | 23.0 | 26.1 | 10.48 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **PE39 M9** | Below Cal | Below Cal | Below Cal | 23.0 | 26.1 | 10.48 |
| **PE41 BB** | Below Cal | Below Cal | Below Cal | 24.0 | 28.9 | 10.48 |
| **PE42 M9** | Below Cal | Below Cal | Below Cal | 24.0 | 28.9 | 10.48 |
| **PE44 BB** | 0.241 | 0.0122 | Below Cal | 24.0 | 28.9 | 6.87 |
| **PE45 M9** | 0.443 | 0.0206 | Below Cal | 24.0 | 28.9 | 6.87 |
| **PE47 BB** | 0.443 | 0.0206 | Below Cal | 25.0 | 28.9 | 6.87 |
| **PE48 M9** | 0.518 | 0.0242 | Below Cal | 25.0 | 28.9 | 6.87 |
| **PE50 BB** | 0.989 | 0.0404 | Below Cal | 30.0 | 28.9 | 6.87 |
| **PE51 M9** | 0.538 | 0.0242 | Below Cal | 30.0 | 28.9 | 6.87 |
| **PE52 BB** | 0.619 | 0.0272 | Below Cal | 30.0 | 28.3 | 6.87 |
| **PE55 BB** | 3.054 | 0.111 | Below Cal | 33.0 | 28.3 | 5.73 |
| **PE56 M9** | 1.216 | 0.048 | Below Cal | 33.0 | 28.3 | 5.73 |
| **PE61 BB** | Below Cal | Below Cal | Below Cal | 33.5 | 27.7 | 5.73 |
| **PE62 M9** | Below Cal | Below Cal | Below Cal | 33.5 | 27.7 | 5.73 |
| **PE64 M9** | Below Cal | Below Cal | Below Cal | 31.0 | 27.2 | 5.73 |
| **PE66 BB** | Below Cal | Below Cal | Below Cal | 31.0 | 27.2 | 5.73 |
| **PE67 BB** | Below Cal | Below Cal | Below Cal | 31.0 | 25.6 | 4.67 |
| **PE68 M9** | Below Cal | Below Cal | Below Cal | 31.0 | 25.6 | 4.67 |
| **PE71 M9** | 0.906 | 0.0329 | 0.032 | 30.0 | 26.7 | 4.67 |
| **PE73 BB** | 0.523 | 0.0194 | 0.019 | 31.0 | 28.3 | 4.67 |
| **PE74 M9** | 1.774 | 0.0631 | 0.06 | 31.0 | 28.3 | 4.67 |
| **PE79 BB** | 2.185 | 0.0754 | 0.075 | 32.0 | 26.1 | 1.91 |
| **PE82 BB** | 0.548 | 0.0196 | 0.02 | 32.0 | 25.6 | 1.91 |
| **PE83 M9** | 0.826 | 0.0301 | 0.029 | 32.0 | 25.6 | 1.91 |
| **PE85 BB** | Below Cal | Below Cal | Below Cal | 35.0 | 25.6 | 1.91 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **PE86 BB** | 0.547 | 0.0199 | 0.02 | 32.0 | 24.4 | 1.91 |
| **PE87 M9** | 0.628 | 0.0224 | 0.0224 | 32.0 | 24.4 | 1.91 |
| **PE89 BB** | 0.699 | 0.0256 | 0.0257 | 31.0 | 25.0 | 1.91 |
| **PE90 M9** | 2.023 | 0.0702 | 0.07 | 31.0 | 25.0 | 1.91 |
| **PE92 BB** | 1.549 | 0.054 | 0.0549 | 28.0 | 25.0 | 1.17 |
| **PE93 M9** | 1.371 | 0.0483 | 0.048 | 28.0 | 25.0 | 1.17 |
| **PE95 M9** | Below Cal | Below Cal | Below Cal | 31.0 | 25.0 | 1.17 |
| **PE96 M9** | Below Cal | Below Cal | Below Cal | 31.0 | 25.0 | 1.17 |
| **PE98 BB** | 0.441 | 0.0239 | 0.024 | 30.0 | 24.4 | 1.17 |
| **PE99 M9** | 0.408 | 0.0229 | 0.023 | 30.0 | 24.4 | 1.17 |
| **PE101 BB** | 0.529 | 0.0266 | Below Cal | 34.0 | 24.4 | 1.17 |
| **PE102 M9** | 0.511 | 0.026 | Below Cal | 34.0 | 24.4 | 1.17 |
| **PE104 BB** | 0.503 | 0.0258 | Below Cal | 33.0 | 25.0 | 2.61 |
| **PE105 M9** | 0.524 | 0.0265 | Below Cal | 33.0 | 25.0 | 2.61 |
| **PE107 BB** | 0.546 | 0.0272 | Below Cal | 34.0 | 25.0 | 2.61 |
| **PE108 M9** | 0.543 | 0.0262 | Below Cal | 34.0 | 25.0 | 2.61 |
| **PE110 BB** | 0.544 | 0.027 | Below Cal | 34.0 | 25.0 | 2.61 |
| **PE111 M9** | 0.542 | 0.027 | Below Cal | 34.0 | 25.0 | 2.61 |
| **PE113 BB** | 0.518 | 0.0262 | Below Cal | 34.0 | 23.9 | 2.61 |
| **PE114 M9** | 0.547 | 0.0271 | Below Cal | 34.0 | 23.9 | 2.61 |
| **PE119 BB** | 0.409 | 0.0231 | Below Cal | 35.0 | 25.0 | 1.18 |
| **PE120 M9** | Below Cal | Below Cal | Below Cal | 35.0 | 25.0 | 1.18 |
| **PE122 BB** | Below Cal | Below Cal | Below Cal | 33.0 | 25.6 | 1.18 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **PE123 M9** | Below Cal | Below Cal | Below Cal | 33.0 | 25.6 | 1.18 |
| **PE125 BB** | Below Cal | Below Cal | Below Cal | 33.0 | 26.1 | 1.18 |
| **PE126 M9** | Below Cal | Below Cal | Below Cal | 33.0 | 26.1 | 1.18 |
| **PE128 BB** | Below Cal | Below Cal | Below Cal | 34.0 | 26.7 | 3.3 |
| **PE129 M9** | Below Cal | Below Cal | Below Cal | 34.0 | 26.7 | 3.3 |
| **PE131 BB** | Below Cal | Below Cal | Below Cal | 33.0 | 26.7 | 3.3 |
| **PE132 M9** | Below Cal | Below Cal | Below Cal | 33.0 | 26.7 | 3.3 |
| **PE134 M9** | Below Cal | Below Cal | Below Cal | 34.0 | 26.7 | 3.3 |
| **PE136 BB** | Below Cal | Below Cal | Below Cal | 34.0 | 26.7 | 3.3 |
| **PE138 M9** | Below Cal | Below Cal | Below Cal | 33.0 | 27.2 | 3.3 |
| **PE140 BB** | Below Cal | Below Cal | Below Cal | 34.0 | 27.2 | 3.3 |
| **PE141 M9** | Below Cal | Below Cal | Below Cal | 34.0 | 27.2 | 3.3 |
| **PE143 BB** | Below Cal | Below Cal | Below Cal | 32.0 | 28.3 | 19.63 |
| **PE144 M9** | Below Cal | Below Cal | Below Cal | 32.0 | 28.3 | 19.63 |
| **PE146 BB** | Below Cal | Below Cal | Below Cal | 33.0 | 28.3 | 19.63 |
| **PE147 M9** | Below Cal | Below Cal | Below Cal | 33.0 | 28.3 | 19.63 |
| **PE149 BB** | Below Cal | Below Cal | Below Cal | 25.0 | 28.9 | 19.63 |
| **PE150 M9** | Below Cal | Below Cal | Below Cal | 25.0 | 28.9 | 19.63 |

APPENDIX 3: Order and family taxonomic stacked bar charts.